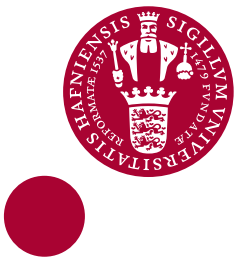


A transcriptomics and systems biology approach to identify candidate genes and biological pathways determining residual feed intake in Danish dairy cattle

Salleh, Suraya Binti Mohamad

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

A transcriptomics and systems biology approach to identify candidate genes and biological pathways determining residual feed intake in Danish dairy cattle

Suraya Mohamad Salleh
PhD Dissertation

UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES
DEPARTMENT OF VETERINARY AND ANIMAL SCIENCES
PHD THESIS 2018 · ISBN 978-87-7209-203-4

SURAYA MOHAMAD SALLEH

A transcriptomics and systems biology approach to identify candidate genes and biological pathways determining residual feed intake in Danish dairy cattle

SL grafik

A transcriptomics and systems biology approach to identify candidate genes and biological pathways determining residual feed intake in Danish dairy cattle

Suraya Mohamad Salleh

PhD Dissertation

Department of Veterinary and Animal Sciences
Faculty of Health and Medical Sciences
University of Copenhagen

This dissertation has been submitted to the
Graduate School of the Faculty of Health and Medical Sciences
University of Copenhagen

Copenhagen, Denmark
2018

Name of department: Department of Veterinary and Animal Sciences

Author: Suraya Binti Mohamad Salleh

Title and subtitle: A transcriptomics and systems biology approach to identify candidate genes and biological pathways determining residual feed intake in Danish dairy cattle

Topic Description: Integrative systems biology analysis using transcriptomics and genomics data to identify candidate genes and biological pathways associated to residual feed intake (RFI) in Danish dairy cows

Supervisor: Mette Olaf Nielsen (Professor)
University of Copenhagen
Department of Veterinary and Animal Sciences
Section of Production, Nutrition and Health
Grønnegårdsvej 3, 1-61
1870 Frederiksberg C, Denmark

Co-supervisors: Haja Kadarmideen (Professor)
Technical University of Denmark
Richard Petersens Plads
Building 324, room 282
2800 Kgs. Lyngby, Denmark

Peter Løvendahl (Senior Scientist)
Aarhus University
Department of Molecular Biology and Genetics - Center for Quantitative Genetics and Genomics
Blichers Allé 20,
8830 Tjele, Denmark

Submitted on: 26th July 2018

Cover Illustration: Holsteins in a barn

A transcriptomics and systems biology approach to identify candidate genes and biological pathways determining residual feed intake in Danish dairy cattle
Ph.D. Dissertation 2018© Suraya Mohamad Salleh (Submitted on July 2018)

ISBN 978-87-7209-203-4
Printed by SL grafik, Frederiksberg, Denmark (www.slgrafik.dk)

Preface

This dissertation is an integral part of the on-going externally funded large research project called "Feed Utilization in Nordic Cattle (FUNC)" at the University of Copenhagen. The work presented in this dissertation was carried out between October 2014 and July 2018 under the supervision of Professor Haja Kadarmideen, Senior Scientist Peter Løvendahl and Professor Mette Olaf Nielsen. The experiment was funded by Danish Milk Levy Foundation, Skejby Denmark, the PhD stipend by Universiti Putra Malaysia and Ministry of Higher Education, Malaysia as well as a tuition fee waiver from the University of Copenhagen.

During this dissertation, I took part and I could learn all the steps involved in the experiment, which are tissue sampling, data collections, bioinformatics analysis and data interpretation.

During the first part of the PhD project, I changed research environment for about six weeks to take part in the sampling processes that were conducted at the Danish Cattle Research Centre (DCRC), Foulum, Aarhus University, Denmark. Towards the end of my PhD period, I spent one year and a half at the Department of Bio and Health Informatics, Technical University of Denmark (DTU), to complete the bioinformatics analysis. The dissertation is based on 3 papers (at the time of submission the first paper is published, one paper has passed first review and the last paper is accepted for publication):

Paper 1: RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. **Salleh, M. S.,** Mazzoni, G., Höglund, J. K., Olijhoek, D. W., Lund, P., Løvendahl, P., & Kadarmideen, H. N. (2017). BMC Genomics, 18, 258. http://doi.org/10.1186/s12864-017-3622-9

Paper 2: Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency. **Salleh, M. S.,** Mazzoni, G., Løvendahl, P., & Kadarmideen, H. N. Submitted to BMC Bioinformatics. (Passed 1st review)

Paper 3: Identification of expression QTLs targeting candidate genes for residual feed intake in Danish dairy cattle using systems genomics. **Salleh, M. S.,** Mazzoni, G., Nielsen, M.O., Løvendahl, P., & Kadarmideen, H. N. Submitted to Journal of Genetics and Genome Research. (In press)

Acknowledgements

In the name of Allah, Most Gracious, Most Merciful….

First, my deep sense of gratitude to my supervisors Prof. Mette Olaf Nielsen, Prof. Haja Kadarmiden and Senior Scientist Peter Løvendahl for their encouragement, valuable suggestions and advice during the entire period of the research, and guidance towards the completion of this PhD. These people are the backbone of this PhD research, without them, I will not be in Denmark at all.

My special thanks and heartfelt appreciation to Ministry of Higher Education, Malaysia and Universiti Putra Malaysia (UPM) for granting me with the scholarship and possibility for me to further my PhD study in Denmark. Deep appreciation is also extended to the Graduate School of Health and Medical Sciences Faculty for accepting my PhD study as well as Danish Milk Levy Foundation, Skejby Denmark for granting the 'FUNC' consortia project which funded the RNA-sequencing analysis.

In addition, thank you very much to my other colleagues and former colleagues that always keep me patient and not to give up in particular Gianluca Mazzoni, Gilda Kischinovsky, Markus Drag, Lisette Kogelman, Sharmila Ahmad, Duy Ngoc Do, Rúta Skinkyté-Juskiené, Victor Carmello. Especially Gianluca, thank you so much for everything, I could not make the list what you have helped me because it is countless. Without you perhaps this PhD will only be a dream. My appreciation to the co-authors of my papers, including Dana Olijhoek, Johanna Höglund and Peter Lund. Thank you for your inputs, suggestions and supports. It has been a great pleasure to work with all of you. Thanks also to all staff who involved in the experimental works (Feeding trial) at the DCRC, Foulum, Aarhus University, Denmark.

Thank you so much also to all staffs and friends at the Section of Production, Nutrition and Health as well as the Section of Anatomy, Biochemistry and Physiology especially to Assoc. Prof. Hanne Helene Hansen, Jeanne Oakman, Anni Christiansen, Anne Friis Petersen and others for the hospitality and helping me a lot during the completion of my PhD study.

My sincere appreciation also goes to Assoc. Prof. Dr. Halimatun, Prof. Jothi Malar Panandam, Dr. Sumita Sugnaseelan and Prof. Loh Teck Chwen and other lecturers, staffs and friends at the Department of Animal Science, Faculty of Agriculture, UPM, for their supports since the day before I started my PhD study. I really appreciate it.

I would also like to acknowledge my Malaysian friends, especially Sharmila, Suziana, Shamrul Azhar, Sazuan Nazrah, Zainatul Bahiyah, Rohana, Nazikussabah, Tamil, Diana, Nur Liyana Akmal and a lot others that not mentioned here, without you guys this journey will not be as sweet as I remember. We have been through the same phase together, we understand each other more than anyone else. I love you all.

My heartfelt appreciation to my mother, my father and my brothers for their continuous supports and understanding.

Above all, my heartfelt thanks to the Almighty Allah Subhanahu wa ta'ala for giving me the knowledge, strength and will power to complete this undertaking.

Suraya Mohamad Salleh
Copenhagen, Denmark 2018

Table of contents

Summary (EN)

Introduction: Feed efficiency (FE) is an important subject in livestock production, because the efficient conversion of feed into animal product has a wide range of impacts, including carbon footprint, food resources and arable land use for livestock feed production as well as social and economic aspects. Feed efficiency in dairy cattle, defined as feed-to-milk ration varies tremendously over the course of lactation due to opposite changes in milk production and body fat mobilization or deposition. Feeding is of vital importance during lactation, as diet alters the efficiency of dairy cows by affecting their energy metabolism. Residual feed intake (RFI), which is the difference between observed and predicted feed intake, has been proposed as an alternative measure of efficiency in dairy cattle, but its use in practice is limited by difficulties relating to measurement of feed intake in group fed cows. There is therefore a need for refinement of methodologies to improve accuracy in genetic selection for a sustainable development of milk production. It was hypothesised that understanding of the biological mechanisms and the role of gene expression patterns across the whole genome (transcriptomics) in relation to feed efficiency and diet will improve the accuracy in identification of the most feed efficient animals.

Objectives, Materials and Methods: The aims of this PhD project were to evaluate whether RNA-sequencing transcriptomics technology and systems biology approaches could be used to identify hitherto unrecognised candidate genes and biological mechanisms information that could be used as biomarkers describing differences in residual feed intake (RFI) between two groups of cows with extremely different RFI (high and low) and of two different breeds (Danish Holstein and Jersey). This included assessment of how the cows responded to different diets with very different forage: concentrate ratio with respect to RFI and gene expression profiles. The distinct patterns of changes in expression profiles were characterized using systems biology and integrative approaches: Differentially Expressed Genes (DEGs) analysis, Weighted Gene Co-expression Network Analysis (WGCNA) and expression QTL (eQTL) mapping.

Results and Discussion: The DEGs analysis identified 70 genes in Holstein and 19 genes in Jersey, which were significantly up or downregulated in high-RFI (i.e. low-FE) cows. In the Jersey breed, the expression level of two genes (*FIZ1* and *SEC24D*) was affected differently by the diet in high- and low-RFI cows. The DEG, which both breeds of cows had in common, were all found in the functional enrichment analysis to be associated with the "primary immunodeficiency" pathway. The WGCNA analysis identified several groups of co-regulated genes, which in a breed specific way were associated with high or low RFI. Only few of these

genes were also represented among the DEGs. Interestingly, the upstream regulators and the functional analysis showed that in Holstein the co-expressed genes were involved in cholesterol and lipid metabolism, while in Jersey the genes encoded for factors involved in immune-related functions. We suggest that the identified hub genes (HGs) (central genes among co-expressed genes) of these modules together with the DEGs could be biologically meaningful candidates to consider for the development of RFI (and hence FE) biomarkers. The integration of genomics and transcriptomics data identified several eQTL regions targeting the RFI candidate genes. Among them, *BDH2*, *CHRNE*, *ELOVL6*, *GIMAP4*, *UHRF1, HSD17B4* and *FDXR* were in genomic loci previously associated with major traits influencing FE. Genetic variants can explain the mechanisms behind the results of previous association studies.

Conclusions: Our findings have provided additional information about genes and co-expression patterns that relate to RFI and hence FE of dairy cows, and interestingly, they suggest that different breeds of cattle may adopt different biological strategies (lipid metabolism versus immune related) to achieve high FE. These candidate genes and breed specific differences obviously need to be confirmed in larger populations of dairy cows to validate their potential use for development of new biomarkers. Provided such validation studies confirm the findings of the present study, the identified SNPs (single nucleotide polymorphisms) could be used to develop new or refine existing genomic selection methods by making use of the biological or functional information on the SNPs.

Sammendrag (DA)

Introduktion: Fodereffektivitet (FE) er et vigtigt emne i den animalske produktion, fordi effektiv omdannelse af foder til animalsk produkt har en bred vifte af konsekvenser, herunder for CO2-aftryk, fødevareressourcer og anvendelse af agerjord til foder produktion, samt sociale og økonomiske forhold. Fodereffektivitet hos malkekvæg, defineret som forbrug af foder i forhold til mængden af produceret mælk, varierer voldsomt i løbet af laktationen på grund af modsatrettede ændringer i mælke produktion og mobilisering eller deponering af kropsfedt. Fodring er af afgørende betydning under laktationen, da diæten kan ædnre malkekøernes effektivitet ved at ændre deres energi metabolisme. Residualt foder indtag (RFI), som er forskellen mellem observeret og prædikteret foderoptagelse, er blevet foreslået som et alternativt mål til malkekvæg, men dets brug i praksis er begrænset af vanskeligheder relateret til måling af foderoptagelse i gruppe-fodrede køer. Der er derfor et behov for at raffinere metoder for at kunne forbedre nøjagtigheden i den genetiske udvælgelse med henblik på en bæredygtig udvikling af mælkeproduktion. Hypotesen var, at forståelse af de biologiske mekanismer og betydning af genekspressions mønstre på tværs af hele genomet (transcriptomics) i forhold til fodereffektivitet og fodersammensætning vil forbedre nøjagtigheden med hensyn til identifikation af de mest fodereffektive dyr.

Formål, materialer og metoder: Formålene med dette ph.d.-projekt var at vurdere, om RNA-sekventering transcriptomics teknologi og systembiologiske tilgange kan anvendes til at identificere hidtil uerkendte kandidatgener og biologiske mekanismer, der kan anvendes som biomarkører, til at beskrive forskellene i residual foderindtag (RFI) mellem to grupper af køer med meget forskellige RFI (høj og lav) og fra to forskellige racer (Dansk Holstein og Jersey). Dette omfattede vurdering af hvordan køerne reagerede på forskellige diæter med meget forskellige kraftfoder:grovfoder forhold med hensyn til RFI og gen ekspressions mønstre. De specifikke ændringer i ekspressions mønstre blev karakteriseret ved anvendelse af systembiologi og integrerede tilgange: Differentiel expression af gen (DEG) analyse, vægtede gen co-ekspressions netværks analyse (WGCNA) og ekspressions QTL (eQTL) kortlægning.

Resultater og diskussion: DEG analysen identificerede 70 gener i Holstein og 19 gener i Jersey, som var betydeligt op- eller nedreguleret i høj-RFI (dvs. lav-FE) køer. I Jersey racen blev ekspressionsniveauet af to gener (*FIZ1* og *SEC24D*) påvirket forskelligt af diæten i høj- sammenlignet med lav-RFI køer. De DEG'er, som begge racer af køer havde til fælles, var alle knyttet til "Primær immundefekt" vejen i den funktionelle berigelses analyse. WGNCA analysen

identificerede flere grupper af co-regulerede gener, som på en race-specifik måde var associeret med høj eller lav RFI. Kun få af disse gener var også repræsenteret blandt DEG'er. Det var interessant, at opstrøms regulatorerne og den funktionelle analyse viste, at de co-udtrykte gener hos Holstein var involveret i lipid og kolesterol metabolisme, mens generne for Jersey kodede for faktorer, der er involveret i immun-relaterede funktioner. På baggrund af dette PhD projekt kan det derfor foreslås, at de identificerede hub gener (HG) (centrale gener blandt sam-udtrykte gener) i disse moduler sammen med DEG'er kunne være biologisk meningsfyldte kandidater man kunne overveje at anvende til at udvikle biomarkører for RFI (og dermed FE). Ved at integrere genom og transcriptom data blev der identificeret flere eQTL regioner som har RFI kandidatgenerne som mål. Blandt dem var *BDH2, CHRNE, ELOVL6, GIMAP4, UHRF1, HSD17B4* og *FDXR* i genomisk loci, der tidligere har været forbundet med væsentlige egenskaber, der påvirker FE. Genetiske varianter kan forklare mekanismerne bag resultaterne i tidligere associationsstudier.

Konklusioner: Resultater af dette projekt har givet yderligere oplysninger om gener og co-ekspressionsmønstre, der vedrører RFI og dermed FE hos malkekøer, og det er interessant at de antyder, at forskellige kvægracer kan antage forskellige biologiske strategier (lipid metabolisme versus immun relaterede) for at opnå høj FE. Disse kandidatgener og race specifikke forskelle skal selvfølgelig skal bekræftes i større populationer af malkekøer for at kunne validere deres potentielle anvendelse til udvikling af nye biomarkører. Forudsat at sådanne valideringsundersøgelser bekræfter resultaterne af det foreliggende studie, vil man kunne anvende de identificerede SNP'er (enkeltnukleotid polymorfismer) til at udvikle nye eller forbedre eksisterende genomiske selektionsmetoder ved at gøre brug af de biologiske eller funktionelle oplysninger i disse SNP'er.

List of abbreviations

ADG: Average daily gain

BCS: Body Condition Score

BLUP: Best Linear Unbiased Prediction

BP: Biological Process

BW: Body Weight

C: Control (Low Concentrate)

CC: Cellular Component

cDNA: complimentary DNA

CEG: Co-expressed Genes

DCRC: Danish Cattle Research Centre

DE: Differentially Expressed

DEGs: Differentially Expressed Genes

DGE: Differential Gene Expression

diceR: Diverse Cluster Ensemble in R

DiffCoEx: Differential Co-expression

DINGO: Differential Network Analysis in Genomics

DM: Dry Matter

DMI: Dry Matter Intake

DNA: deoxyribonucleic acids

ECM: Energy Corrected Milk

eQTL: expression Quantitative Trait Loci

FCE: Feed Conversion Efficiency

FCR: Feed Conversion Ratio

FDR: False Discovery Rate

FE: Feed Efficiency

GLM: Generalized Linear Model

GO: Gene Ontology

GS: Gene Significance

GSEA: Gene Set Enrichment Analysis

GWAS: Genome Wide Association Studies

HC: High Concentrate

HGs: Hub Genes

HWE: Hardy Weinberg Equilibrium

IPA: Ingenuity Pathway Analysis

KEGG: Kyoto Encyclopedia of Genes and Genomes

KR: Kleiber Ratio

MAF: Minor Allele frequency

ME: Module Eigengene

MF: Molecular Function

miRNAs: micro RNAs

MM: Module Membership

mRNA: messenger RNA

MS: Milk Solids

MTR: Module Trait Relationship

NEB: Negative Energy Balance

NRC: National Research Council

padj: Adjusted P-value

PCA: Principal Component Analysis

PCCs: Pearson Correlation coefficients

PPI: Protein-protein Interaction

QTL: Quantitative Trait Loci

RFI: Residual Feed Intake

RIN: RNA Integrity Number

RLE: Relative Log Expression

RNA: Ribonucleic acid

RNAse: Ribonuclease

RNA-seq: RNA sequencing

RSP: Residual Solids Production

RT-qPCR: Real-Time quantitative Polymerase Chain Reaction

SNP: Single Nucleotide Polymorphism

STAR: Spliced Transcripts Alignment to a Reference

STRING: Search Tool for the Retrieval of Interacting Genes/Proteins

TMM: Trimmed Mean of *M*-values

TOM: Topological Overlapped Matrix

WGCNA: Weighted Gene Co-expression Network Analysis

Background of the study

Finding the best animals to be used for breeding is an important foundation for a sustainable livestock production. Genetic markers for many complex traits and disease have been identified by several methods and technologies, and such genetic markers can help and facilitate animal breeders by improving prediction of genetic merit to ensure selection of the best animals for breeding purposes. However, genetic markers are not always accurate and they do not always allow accurate prediction of the production trait of interest. Possible reasons could be that important genetic markers are overlooked; some of the genetic markers could be breed specific and not accounting for physiological differences in other breeds; or the markers could be specific for certain (feeding) conditions.

Almost 30 years ago Adams et al. (1991), when transcriptomics technologies became available, they were seen as promising tools for identification of overlooked candidate genes and development of new biomarkers, as they could help us to understand the genes and biological mechanisms involved in complex traits and diseases. Within transcriptomics, RNA-sequencing technologies have since then been refined and become cheaper, and deep-sequencing technologies can now help us to discover more information about the links between the genome, gene expression profiles (Z. Wang et al., 2009) and desired animal phenotypes.

In this PhD project, it was attempted to combine advanced omics technologies with new systems biology approaches in order to identify candidate genes and biomarkers that potentially could be used for genomic selection in dairy cows for a complex trait like feed efficiency (FE). Feed intake is a major determining factor in dairy cattle for FE, which is an economically important trait in livestock production, and no accurate breeding index/indices exist for FE in dairy cattle, due to a complex underlying biology involving many different body tissues and vast changes in their function over time. Hence for no other livestock production do correlations between phenotypic traits change over time to the extent observed for FE in dairy cattle, where correlations between feed intake and milk yield or cow body weight can change from highly positive to highly negative or vice versa during the course of lactation (Hoffman et al., 2000). Transcriptomics and new systems biology approaches can potentially help us to identify important and hitherto overlooked biomarkers and improve the genetic prediction of the complex traits, feed intake and FE in dairy cattle.

FE represents the ability of an animal to convert ingested feed into animal product, and it is often expressed as the feed conversion ratio (FCR), i.e. the ratio between feed (or dry matter (DM))

intake per unit of animal product. FCR does not accurately take animal size and growth into account, and it is therefore a difficult term to utilize in animal breeding. In recognition of this, Koch et al. (1963) introduced a new term for beef cattle, namely residual feed intake (RFI), which is the difference between observed feed intake and a predicted feed (or energy) intake. The predicted feed or energy intake is calculated based on successive regressions between different phenotypic manifestations, such as live weight, daily weight change and milk production. The RFI has been widely used in poultry, pigs, and beef cattle. It is also a promising tool as a measure of FE in dairy cattle breeding (Berry & Crowley, 2013; E. Connor, 2015; Pryce et al., 2015; Tempelman et al., 2015; VandeHaar et al., 2016; Veerkamp et al., 1995), although the problems of accurate genetic merit for feed intake prediction are well recognised due to the complex underlying biology with enormous shifts in correlations between signifying phenotypic traits in addition to problems of accurate determination of actual individual feed intake in group housed cows.

Several Genome Wide Association Studies (GWAS) have focused on FE or RFI in beef cattle (Santana et al., 2014; Saatchi et al., 2014), but attempts to accurately predict FE in dairy cattle by this approach has not been very successful (Berry & Crowley, 2013). Integration of genomics with transcriptomics data and subsequent application of systems biology approaches to identify expression Quantitative Trait Loci (eQTL) has been used to more accurately identify important genomic loci and candidate genes for FE in other species, and can indirectly provide information about SNPs/eQTLs associated with FE.

There are a small number of studies, where omics and systems biology approaches have been applied in relation to milk production and fertility (Bu et al., 2017; Dai et al., 2017; Moran et al., 2016), however, transcriptomics, eQTL and systems biology approaches have never been applied in any studies with dairy cattle to our knowledge to assess their utility to improve prediction of complex traits such as FE or RFI.

The present PhD was based on the hypothesis that application of transcriptomics and new systems biology analyses can unravel hitherto overlooked candidate genes and identify new biomarkers responsible for genetic differences for complex biological traits, such as RFI and hence FE in dairy cattle. It was chosen to work with RFI as the model multifactorial trait in this dissertation, since it was possible to obtain biological samples for genomics and transcriptomics analyses from an experiment, where very different RFI had been determined in dairy cows of two different breeds exposed to two very different diets.

Most of the transcriptomic studies relating to FE have used muscle tissues (beef cattle) or mammary tissue or milk (dairy cattle). In the present PhD project, liver tissue was obtained, and

this is another interesting target tissue to study in relation to dairy cattle, since the liver is a crucial organ responsible for metabolic and endocrine adaptations during lactation, it is implicated in the development of production related diseases such as ketosis and fatty liver, and it is responsible for the first line acute phase immune response to e.g. mammary infections (Minuti et al., 2015).

To elucidate the above-mentioned hypothesis, the scientific literature will in the following be reviewed to establish how the new omics and systems biology techniques can be used to complement more conventional genomics studies for complex biological traits, such as RFI and hence FE, and to establish which genetic markers have been identified in studies with other livestock species and beef cattle breeds.

Thereafter, results will be presented from in-depth transcriptomics and integrative systems biology analysis, where the aim was to identify potential candidate genes relating to the model trait RFI in dairy cows for the first time. In the attempt to improve our understanding of the biology underlying the RFI trait, RNA-sequence analyses were applied to liver tissues obtained from lactating cows of different breeds (Danish Holstein and Danish Jersey) and during two experimental periods, when they were fed diets with different forage:concentrate ratios.

The findings of the experimental part of the PhD project have been reported in three different scientific papers:

1) **Paper 1:** Where the aim was to identify differentially expressed genes and biological pathway associated with high- or low-RFI and to evaluate the effects of different diets interacting with RFI level.

   The specific hypothesis for this part of the study was that liver gene expression profiles can reflect differences between high- and low-RFI animals and that the up and down regulated genes are involved in biological mechanisms and pathways regulating RFI.

2) **Paper 2:** Where the aim was to 1) identify groups of co-expressed genes and biological pathways associated with RFI, 2) reveal potential candidate genes (hub-genes and upstream regulators) for RFI-related traits, and 3) compare the mechanisms and processes underlying differences in the RFI trait between Holstein and Jersey cattle.

   The specific hypothesis for this part of the study was that genes, which are up and down regulated together across samples, interact and participate in the same biological mechanisms associated with the RFI.

3) **Paper 3:** Where the aim was to identify eQTL regions targeting candidate genes for the RFI related traits in Holstein and Jersey cattle.

   The specific hypothesis for this part of the study (integrative genomics analysis) was that SNPs associated with the expression of candidate genes are involved in, or in linkage with, genomic regions regulating their expression.

To the best of our knowledge, this is the first study of its kind in dairy cattle, and as described later, candidate genes for RFI were identified that potentially can be used to develop biomarkers for RFI, and hence FE, in dairy cattle. These biomarkers were found to be breed specific and their expression depended to some extent on the dietary exposure (high or low forage:concentrate ratio).

Literature review

Globally, livestock production is really important to fulfil human nutritional needs. There are enormous research efforts and development activities targeting sustainability of livestock production, not least within breeding programs. Numerous technologies and tools have been applied in order to facilitate breeding programs. For instance, omics technologies, such as genomics and transcriptomics, as well as systems biology approaches have been used in order to understand the genetic foundation for the animal's biology to improve the precision in animal breeding.

Omics technologies can be used to improve our understanding of how the genome is expressed in a living organism and related to specific phenotypic traits (Debnath et al., 2010). Thus, the availability of omics technologies and high throughput data has given us opportunities to investigate the mechanisms and the biology underlying complex traits and diseases (Suravajhala et al., 2016). From an animal production perspective, it has been demonstrated in studies with different livestock species and cattle breeds that deeper understanding of the link between the animal genome may help us to improve the prediction of genetic merit for specific performance traits of economic importance. The following will provide a short introduction to key omics technologies and how to extract new information on the function of the animal genome from the bioinformatics pipeline.

Transcriptomics, RNA-sequencing and the bioinformatics pipeline

Transcriptomic (RNA) analyses reveals the level of expression of genes in a specific cell or tissue at a specific time, and in a specific state (Lowe et al., 2017). Transcriptomics has been widely used to study RNA expressions in any relevant tissue. Often, researchers have compared the expression of mRNA or total RNA between two or more groups of animal or correspondent samples. This would allow researchers to investigate the profiles and the differences of the expression relating to different conditions (Alexandre et al., 2015; Bionaz & Loor, 2012; Lee et al., 2015; Paradis et al., 2015). Over the past few years, researchers started using RNA-sequencing (RNA-seq) methods intensively to investigate their RNA specimens (Ramayo-Caldas et al., 2018; Weber et al., 2016; Wickramasinghe et al., 2014; Yi et al., 2015; Z. H. Zhang et al., 2014). This method (RNA-seq) was also applied in the present PhD study.

RNA-seq consists of several steps: RNA isolation, library preparation and sequencing (Kukurba & Montgomery, 2015). The sequencing reaction occurs through sequencing by synthesis cycles, where each nucleotide is added to a growing template and a laser reveals which nucleotides are

added at each cycle. This process allow that the sequencing of millions of fragments can be parallelized, reducing cost and time (Buermans & Den Dunnen, 2014). From RNA-seq, the output is the sequenced reads that can be analysed with a specific bioinformatics pipeline, which will be addressed in a subsequent section.

Next, reads are aligned to the reference genome to generate a comprehensive genome-wide picture of the expression profiles in a tissue, which gives information about, which genes are actively transcribed. This process is usually done by using a bioinformatics computational tool, such as STAR (Dobin et al., 2013), TopHat (Trapnell et al., 2009), GSNAP (Wu & Nacu, 2010), MapSplice (K. Wang et al., 2010) and RUM (Grant et al., 2011).

After sequence alignment, the mapped reads or the gene expressions are measured as gene counts (Dündar et al., 2015). The gene counts for each annotated genes are performed by counting the number of read pairs that overlap the genomic region, where the gene is annotated. The gene expression quantification are commonly done using other computational tools, such as HTSeq (Anders et al., 2014), Cufflinks (Trapnell et al., 2010) and FluxCapacitor (Griebel et al., 2012).

In addition, quality control should be performed in order to identify and correct for possible biases that arise during the RNA-seq pipeline (RNA extraction, sample preparation, library construction and sequencing) (Kukurba & Montgomery, 2015).

Finally, before further gene expression analysis, the quantified gene expressions need to undergo a normalization process to remove biases and to obtain accurate results (Maza, 2016; Mazzoni & Kadarmideen, 2016; Z. Wang et al., 2009). This is important in order to compare the samples expression level at similar gene count levels. The normalization is done by several commonly used computational methods (e.g. RLE (Relative Log Expression) from DESeq2 package (Love et al., 2014), voom from limma (Linear Models for Microarray and RNA-Seq Data) package (Ritchie et al., 2015), TMM (trimmed mean of *M*-values) from edgeR (Empirical Analysis of Digital Gene Expression Data in R) package (Robinson et al., 2010).

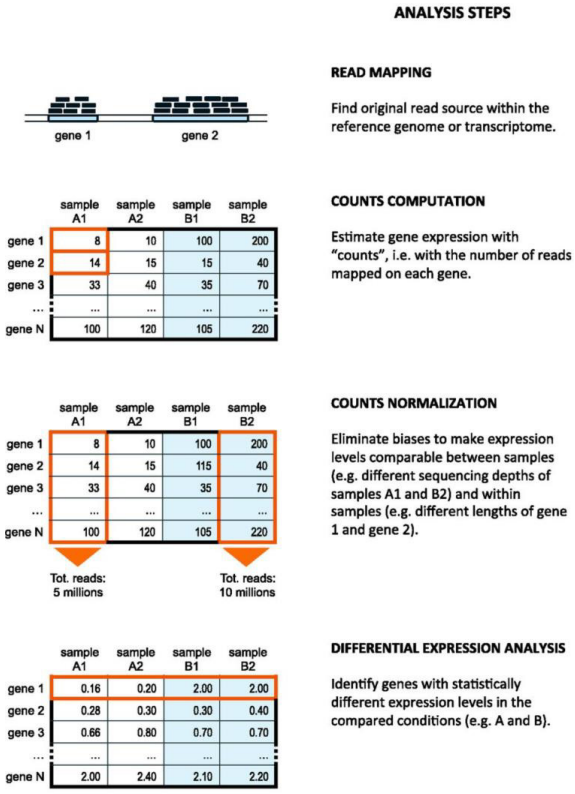
The final output is a matrix of gene expression counts (number of read pairs mapping in the genomic location of an annotated gene) normalized to account for differences in library sizes across samples. Generally, each column in the matrix corresponds to a sample and each row corresponds to one of the annotated genes.

Differential Gene Expression (DGE) analysis

Based on the outputs from the transcriptomics analysis, the differential pattern of expression of individual genes can be analysed further in several different ways. One of the ways is to compare

similarities and differences between the expression of distinct genes across several samples or conditions (e.g. control vs treated; healthy vs diseased) (Finotello & Di Camillo, 2015). Such a Differential Gene Expression (DGE) analysis can help to identify specific genes that play major roles in controlling certain phenotypes. In other words, the analysis will identify differentially expressed genes (DEGs) in the gene expression profile comparison.

The DGE analysis can be done using several types of computational tools and packages that measure comparatively and statistically across groups of samples/conditions. Among the common packages used for the DGE analysis are: limma (Ritchie et al., 2015), edgeR (Robinson et al., 2010), DESeq2 (Love et al., 2014). The DESeq2 assumes a negative binomial distribution of the gene counts and applies shrinkage to the fold changes, and is useful to account for biases due to low expression of certain genes (Love et al., 2014). Figure 1 provides an overview of the RNA-seq pipeline from the read mapping (transcriptomics) step until the DGE analysis.



	sample A1	sample A2	sample B1	sample B2
gene 1	8	10	100	200
gene 2	14	15	15	40
gene 3	33	40	35	70
...	...	...	...	...
gene N	100	120	105	220

	sample A1	sample A2	sample B1	sample B2
gene 1	8	10	100	200
gene 2	14	15	115	40
gene 3	33	40	35	70
...	...	...	...	...
gene N	100	120	105	220

	sample A1	sample A2	sample B1	sample B2
gene 1	0.16	0.20	2.00	2.00
gene 2	0.28	0.30	0.30	0.40
gene 3	0.66	0.80	0.70	0.70
...	...	...	...	...
gene N	2.00	2.40	2.10	2.20

Figure 1 Procedure for differential expression of genes analysis from RNA-seq data. Diagram adapted from Finotello and Di Camillo (2015).

Systems Biology: Weighted gene co-expression network analysis (WGCNA)

After investigating the gene expression profile at the individual gene level (DGE analyses), another common method to understand the gene expression profiles from the RNA-seq technology is to study co-expression patterns of genes to understand not least the expression biology at the system level for multifactorial traits (Kitano, 2002). This is a holistic approach to study and elucidate the biology of complex systems (Breitling, 2010). Systems biology approaches can be used to analyse the entire set of interactions between different molecules, such as RNA, and integrate different biological layers together. Several of the methods used in systems biology analyses are based on networks. Network approaches have been successfully applied to co-expression analysis. In co-expression network analysis, the gene pairwise-correlations are studied to identify groups of genes (modules) that are up or downregulated together across different samples or conditions. It is possible to use the average expression of these modules to select co-expressed genes correlated with a trait of interest. Therefore, the gene co-expression network analysis is also contributing to a better understanding of the biological mechanisms associated with a particular trait (Ram et al., 2012).

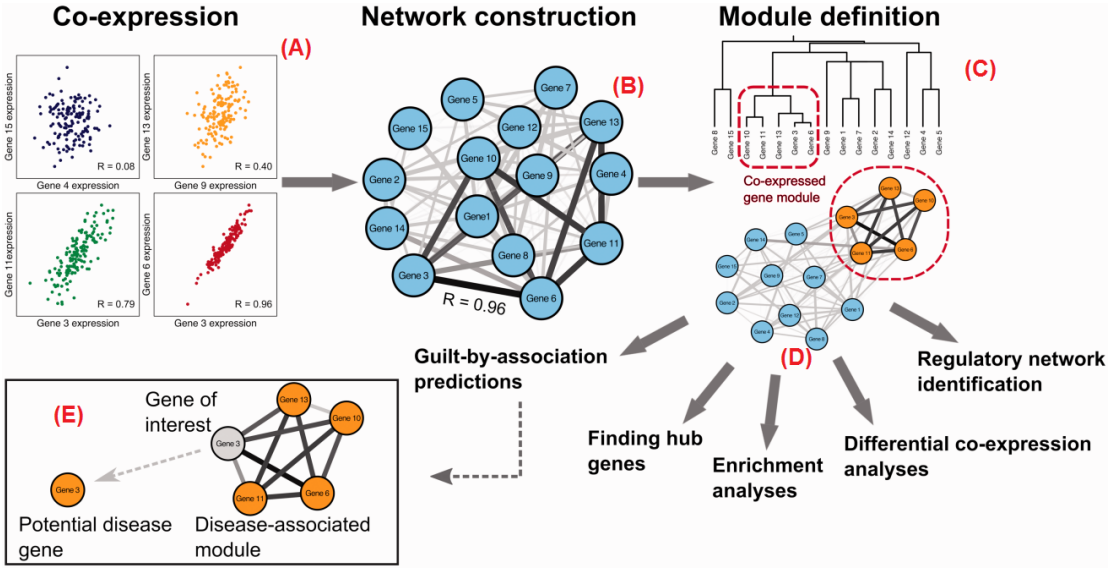


Figure 2 An overview of the flow in a gene co-expression network analysis. Picture from van Dam et al. (2017).

(A) Firstly, a pair-wise correlations is determines for possible gene pair in the gene expression data. (B) Subsequently, all the pair-wise is assigned as a networks. (C) Modules (group of genes) within the networks and then defined by using clustering method. (D) Different type of analysis can be performed to identify regulatory genes, functional enrichment and hub genes. (E) Hub genes identification (genes that have high connectivity with other genes.)

Figure 2 provides an overview of the flow in a co-expression network analysis. By selecting genes with strong intramodular connectivity, it is possible to identify key genes (hub genes/regulators) (Horvath & Langfelder, 2011), which could also be potential candidate genes for the development of biomarkers for the trait of interest. Genes (hub genes) that have high connectivity with other genes are central genes in the modules and could be potential candidate genes for the trait of interest (van Dam et al., 2017).

Several types of tools and packages have been developed for the identification of co-expression modules, such as weighted gene co-expression network analysis (WGCNA) (Langfelder & Horvath, 2008), DICER (Amar et al., 2013), DiffCoEx (Tesson et al., 2010), CoXpress (Watson, 2006), DINGO (Ha et al., 2015), GSCNA (Rahmatallah et al., 2013), GSVD (Alter et al., 2003) and Biclustring (Pontes et al., 2015).

The WGCNA was the choice of methodology in this PhD project, because the weighted analysis produced more robust results compare to unweighted analysis (B. Zhang & Horvath, 2005). It does not set a threshold to the correlations, as it is a weighted approach (soft thresholding). WGCNA is based on scale-free topology assumption of the network (W. Zhao et al., 2010). Firstly, all pair-wise correlations among genes are computed. Next, the beta power of the correlations is computed to meet the scale free assumption. Co-expression network in WGCNA is determined by the interconnectedness level among genes. The beta-power of the correlation is used to compute the Topological Overlap Matrix (TOM) as a measure of interconnectedness between genes. The information of TOM is used to cluster the genes by a hierarchical method. The first component of the gene expression of each module (Eigengene) is used as average representation of the gene expression. Modules whose Eigengene is correlated with a specific trait are further analysed. When a module is strongly associated with a trait, then Gene Significance (GS) and Module Membership (MM) are correlated. Therefore, MM can be used as a parameter to identify hub genes (HGs) that are potential candidate genes for the trait of interest.

Functional enrichment

In addition to understanding of the functions of individual genes and knowledge about how different genes are co-expressed, it can be relevant to obtain a holistic understanding of specific biological mechanism encoded or affected by different sets of genes. This is done in functional enrichment analyses, and by this procedure, it can be established, whether there are groups of genes (among and across those identified in the DGE and WGCNA analyses) that are connected in regulation of a specific (multifactorial) biological function (Hung, 2013). The functional enrichment analysis identifies statistically significant gene sets that represent functions,

mechanisms and processes among the gene set of interest (Hung, 2013). Protein functions are described by the Gene Ontology (GO) terms (molecular function, cellular component and biological process). Information about the pathways is stored in databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000). The functional enrichment can be applied to a set of genes of interest (DEGs, CEGs or other candidate genes) and by including them in a functional enrichment analysis tool, such as GOseq (GO analysis on RNA-seq data) package in R, ClueGO (Young et al., 2010) (an application plugin for Cytoscape software), Search Tool for Retrieval of Interacting Genes (STRING) (Franceschini et al., 2012) database, Gene Set Enrichment Analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005), Ingenuity® Pathway Analysis (IPA®) (Krämer et al., 2013).

Thus, enrichment procedures start by mapping genes (or proteins) according to their biological annotations that can be retrieved from different databases (e.g. GO databases, KEGG, STRING database). Then, the list of target genes (DEGs, CEGs or other candidate genes) is compared against a background (all annotated genes) in order to get significance values for over-represented functional classes of genes/proteins in a set of genes. Hence, the enrichment procedure will identify, which GO terms and biological pathways are statistically over-represented eg. for a specific trait, within the list of gene of interest (Huang et al., 2008; Tipney & Hunter, 2010).

Integrative genomics: Data integration between transcriptomics and genomics

Another alternative approach to identify candidate genes and biomarkers for certain traits is by integrative genomics or genetical genomics approach (Hubner et al., 2005; Le Mignon et al., 2009; Schadt et al., 2005). The genome of an animal is defined at conception and is not tissue specific, but by comparing transcriptomics and genomics data it is possible to get information about any genetic determinants or modifications responsible for specific transcriptomics patterns. Thus, integration of information from transcriptomics and genomics analyses can be used to validate identified candidate genes.

The genetic set-up of the animal can be analyzed by different genotyping techniques, either characterizing the exact nucleotide sequence of every DNA strand, or more commonly by searching for so-called SNPs, which are genes, for which there are single nucleotide polymorphism, i.e. an exchange of one nucleotide from the reference genome for that animal species. SNPs can be located both within gene encoded regions of the DNA or within intergenic regions. The identification of SNPs can be performed by DNA sequencing or by using

genotyping technologies. The high-throughput genotyping technologies identify genotypes of hundreds of thousands SNPs in one single run.

Identification of expression quantitative trait loci (eQTL)

An example of integrative systems genomics analysis is the eQTL mapping. By eQTL mapping, genomic region (SNPs), which correlate with and are assumed to impact (decrease or increase) expression levels of a specific gene, can be identified (Mazzoni et al., 2015). The main assumption is that the regulatory variants (eQTLs) are in linkage with and responsible for up- or downregulating expression of the targeted gene(s). An eQTL can be located within the same loci as the target gene (usually within a one Megabase (Mb) window), termed a *cis* acting or local eQTL, or it can be a *trans* acting or distant eQTL, when it is located far from the loci of the target gene (even at a different chromosome) (Nica & Dermitzakis, 2013). It has long been known that *cis*-regulatory mutations can induce significant changes in animal morphology, physiology and behaviour, and some evolutionary phenotypic changes are more likely to have resulted from *cis*-regulatory mutations than from actual coding mutations (Wray, 2007). It is therefore possible that more identification of eQTLs could provide important new information about the genetics underlying RFI in dairy cattle.

Use of the new omics and bioinformatics technologies to identify candidate genes for feed efficiency

Only very few transcriptomics and bioinformatics studies have been conducted in dairy cattle, and none of these have addressed factors involved in regulation of RFI, FE or feed intake, presumably due to the complicated underlying biology as mentioned previously. There are, however, a number of other recent studies in other livestock species, where the genetics underlying FE has been in focus in the attempt to identify candidate genes and biomarkers for FE related traits. In the following, results from such transcriptomic studies in other livestock species will be presented, and it will be evaluated if the application of these rather expensive technologies can be expected to provide significant improvements to genetic prediction and identification of candidate genes and biological processes associated with FE related traits.

In pig: Many transcriptomic studies have been conducted in pigs relating to FE traits, and they have overall pointed to the utility of transcriptomics studies to identify important predictors for RFI in pigs and to reveal underlying biological pathways. A selected number of studies will be presented here.

The first study that compared transcriptional profiles of high RFI and low RFI pigs was performed by Lkhagvadorj et al. (2009). In that study, RFI was linked to altered expression of genes involved in lipid metabolic processes in adipose tissue in response to caloric restriction.

In another study of the skeletal muscle transcriptome, it was found that 99 mRNAs and 25 miRNAs that were DE and involved in lipid metabolism and biosynthetic process as well as muscle cell growth differentiation were related to RFI (Jing et al., 2015).

In an integrative approach using liver and duodenum tissues, it was discovered that the underlying biology for differences in RFI was related to overrepresented pathways in both tissues involved in oxidative stress response, inflammation and immune response (Ramayo-Caldas et al., 2018). Similarly, in a study conducted with four different tissues (muscle, liver, perirenal fat, subcutaneous fat), it was discovered that the RFI trait was related to biological pathways regulating immune response, protein metabolism and response to oxidative stress in growing pigs (Gondret et al., 2017).

In poultry: Yi et al. (2015) identified 41 DEG in the duodenal transcriptomics analysis of RFI-divergent chicken. The DEGs identified were mainly involved in regulation of digestibility, metabolism, stress response, and energy homeostasis. From a microarray transcriptomic analysis comparing FE in meat type chicken, it was reported that differences between RFI groups could be ascribed to differences in cell division, growth, proliferation and apoptosis, protein synthesis, lipid metabolism, and molecular transport of cellular molecular (Lee et al., 2015). In addition, an integrative analysis of GWAS and transcriptomics of high and low-RFI chicken, discovered regions associated to lipogenesis, social behaviour, and immunity (Xu et al., 2016).

In beef cattle: Transcriptomics studies relating to FE have so far only been conducted in beef cattle. Paradis et al. (2015) identified seven DEG in liver involved in inflammatory processes that were associated with FE trait in beef cattle. In another study, Alexandre et al. (2015) found that gene modules, strongly associated with a low FE trait, were mainly enriched for inflammatory and immune related functions and lipid metabolism. However, Tizioto et al. (2016) found that the main muscle gene differences between feed efficient and inefficient Nellore steers were related to oxidative stress.

In dairy cattle: Transcriptomics studies is a relatively new discipline in dairy cattle, and such studies have hilherto focused only on expression of genes related to milk protein synthesis, fertility and metabolic diseases such as ketosis and milk fever (Bionaz & Loor, 2012; Loor et al., 2007), and none are relating directly to FE traits. There is one study in dairy cattle, which aimed to find DEG in livers from divergently mild as compared to severely negative energy balance (NEB) dairy cattle (McCabe et al., 2012). The pathways related to the severe NEB in that study

were related to regulation of fat metabolism and steroid hormone biosynthesis. Pathways that are associated to NEB could also be key factors influencing FE in dairy cattle, but more extensive studies are obviously needed, where omics technologies and the new systems biology approaches are combined to improve our understanding of the complex genetics and biology underlying the FE trait in dairy cattle.

Based on the outcomes from the previous studies across different livestock species and tissues, it is tempting to assume that the omics technologies and new systems biology approaches could be valuable to apply also in dairy cattle breeding to improve genetic prediction for an economically important multifactorial trait as FE. This has to the best of the author's knowledge never previously been done. In this dissertation, the aim was to evaluate the utility of these omics technologies combined with new systems biology approaches to identify candidate genes that can be used for the development of biomarkers that determine RFI. Feed intake is a very important factor for feed efficiency, and it was a previously mentioned, chosen as the model multifactorial trait to work with on this PhD project, based on biological samples available from an experiment, where very different residual feed intakes (RFI) had been determined in dairy cows of two different breeds exposed to two different diets. In the following, RFI and the importance of FE will therefore briefly be introduced.

Importance of feed efficiency in dairy production

Livestock production plays a major role to feed the global populations. Milk and milk products have been consumed by populations worldwide and are important sources of protein, carbohydrates, vitamins, mineral and fat (Wong, 2012). In Northern Europe, Denmark is one of the top producers of dairy products. There are consequences associated with dairy production that we need to be aware of, such as the use of arable land for animal feed production, contribution of rumen fermentation to greenhouse gas emission, global warming and other emissions from animal waste (Bilotta et al., 2007; Knapp et al., 2014; Steinfeld et al., 2006; White & Hall, 2017). There are several elements that need to be considered by every sector (government, private sectors, farmers and consumers) in order to ensure a sustainable development of livestock production to meet future consumer demands. From the animal production as well as breeding perspective, one crucial thing to consider in livestock production is FE, namely to convert feed into produce efficiently, and in a sustainable way.

Achieving reductions in feed consumption, while maintaining milk production, will improve FE as well as the farmer revenue. To achieve this type of animal for high profit productions, both nutritional and genetic means for improvements need to be considered.

In livestock production, feed represents more than 50 per cent of the total production costs (E. E. Connor et al., 2012; Lawrence et al., 2008). FE therefore not only plays an important role for the production economy, but can also contribute to reduce the environmental foot print of production by reducing greenhouse gas emission and animal waste per unit of produce (milk and meat) (Madsen et al., 2010).

Feed efficiency (FE) measurements in dairy cattle

In general, FE in other livestock system such as poultry, pig, beef cattle (meat purpose animals) as well as dairy cattle production has been studied for years (Gilmore, 1952; Wilkinson, 2011). Measuring FE in lactating animals is more difficult compared to measuring FE in growing animals, because of the complex biology underlying lactation, and FE in dairy cows is affected by many non-genetic factors, which complicate the development of accurate breeding indices for FE. The major non-genetic factors are related to feeding (DMI, type of feed, forage digestibility etc.) (Casper, 2008; Laflamme, 1973) and current physiology of the cow (energy balance, milk production, activity, reproduction etc.) (Allen & Bradford, 2009). Dairy cows need to consume feed to synthesize milk and to generate energy and provide nutrients to maintain body reserves during the lactation and dry periods (Council, 2001). However, immediately after calving, lactating animals normally undergo rapid catabolism of body reserves due to an insufficient feed intake capacity relative to nutrient output in milk, and this is followed by anabolism of body reserves in later stages of lactation and until next calving (Berry & Crowley, 2013; Roche et al., 2009). Thereby the feed-to-milk conversion ratio changes markedly over time for a given cow, irrespectively of her genetic merit.

Berry and Crowley (2013) have therefore reviewed genetic correlations between a number of different phenotypic traits relating to FE in dairy and beef cattle. A large proportion of the genetic variation in feed intake could be ascribed to five predictor traits: body weight (BW), growth rate, milk yield, body composition, and linear type traits reflection e.g. body size in both types of cattle. Table 1 provides an example of some important traits that have been used to define differences in FE in dairy cattle.

Table 1 **Different definitions of FE used for dairy cattle**

Trait name	Abb.	Definition	Formula	Source
Feed Conversion Efficiency	FCE	Efficiency of feed being converted directly to milk solids or ECM*	kg of milk solids/kg of total DMI	(Coleman et al., 2010)
Milk solids per kg of BW (milk production potential)	MS	Milk solids or ECM produced per kg of BW	kg of milk solids/kg of BW	(Coleman et al., 2010)
Dry Matter Intake per kg of BW (feed intake capacity)	DMI	Total Dry Matter Intake per KG of BW	Total DMI/kg of BW	(Coleman et al., 2010)
Residual Feed Intake (feed intake deviation from predicted)	RFI	FI net of the expected feed requirements for maintenance and growth, with expFI obtained by regression	total DMI_t − ($year_y$ + fat $yield_t$ + protein $yield_t$ + lactose $yield_t$ + $BW_t^{0.75}$ + ΔBW_t + BCS_t)	(Koch et al., 1963)
Residual solids production (milk production deviation from predicted)	RSP_t	Residual solids production may be defined as the actual milk solids produced relative to expected solids production based on the feed intake of an individual animal and other energy sinks (e.g., maintenance, growth) or energy sources (e.g., body tissue mobilization).	milk solids $yield_t$ − ($year_y$ + total DMI_t + $BW_t^{0.75}$ + ΔBW_t + BCS_t)	(Coleman et al., 2010)

* ECM: Energy Corrected Milk

Feed intake is obviously a major determinant of FE, but the actual feed intake of a cow correlates with many factors, such as the volume of the gastrointestinal system and hence animal size, of BW and body conformation (stature) in different stages of lactation and may also depend on the breed. Extent of fatness and hence Body Condition Score (BCS) can influence the extent of body fat mobilization as well as feed intake of the cow in early lactation.

Residual Feed Intake (RFI) is a measure of the difference between actual and predicted feed intake, and it was proposed as a useful trait in breeding programs to describe feed efficiency (E. E. Connor et al., 2012), since it can take challenging changes during the course of lactation into

account. This term was therefore used as a model trait in the present PhD project, and will be described further in the following.

RFI as a measure for FE in dairy cattle and the physiological basis

In recognition of the difficulties involved with the use of FE in dairy cattle breeding, an alternative term, Residual Feed Intake (RFI) was developed and proposed for beef cattle by Koch et al., (1963). The RFI is computed as the difference between actual and predicted feed intake (Koch et al., 1963), and has the advantage of being independent of e.g. an animal's mature size and growth rate (Moore et al., 2009). The predicted feed intake is calculated as the residual in a linear model that include several covariates between feed intake and animal performance characteristics (milk energy, metabolic body weight, body weight change and cohort effect) (Potts et al., 2015), see Figure 3. The lower the RFI values, the more feed efficient is the animal. The estimation of RFI in lactating cows rely on linear models establishing correlations between important factors/traits, which include test period duration, stage of lactation, parity number, milking frequency, frequency of nutrient analysis, predicted daily BW, daily energy corrected milk (ECM) yield as well as several others, as described in detail in (E. E. Connor et al., 2012).

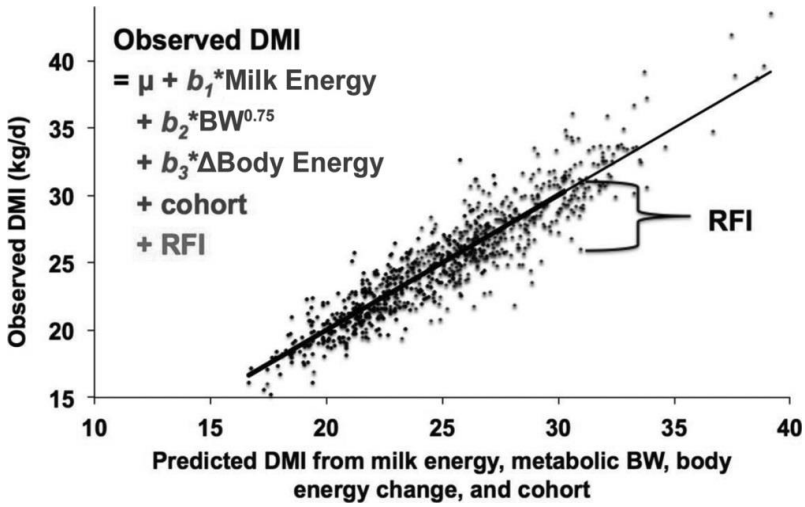


Figure 3 An example of a plot of the distributions of observed versus predicted feed intake and the determination of (RFI) as the difference between the two. Cows that that eat less than the predicted feed intake are more efficient in converting feed gross energy to net energy or require less net energy for maintenance than expected based on their BW. Graph from VandeHaar et al. (2016).

Due to lack of accurate measurements of feed intake for individual cows, the use of RFI is not quite as established in dairy production as in other livestock productions, and the use of RFI is complicated by the lack of precision of the estimate. Moreover, the utility of RFI in dairy cows

as a selection criteria, may be complicated by the potential negative associated effects in the form of poor reproduction performance and more pronounced loss of body weight in early lactation (Pryce et al., 2012), if the complex underlying biology is not properly accounted for. Feed intake in dairy cattle is determined by interactions between many different organ systems in the body of the animal, which give rise to dramatic changes in correlations between milk production, DMI and live weight change over the stage/week of lactation along with the transition from negative to positive energy balance (see Figure 4). Bingjie Li et al. (2017) therefore suggested that the weeks of lactation should be accounted for, when assessing RFI in lactating cows, to account for the changes in regression coefficients between DMI on energy sinks over the course of lactation.

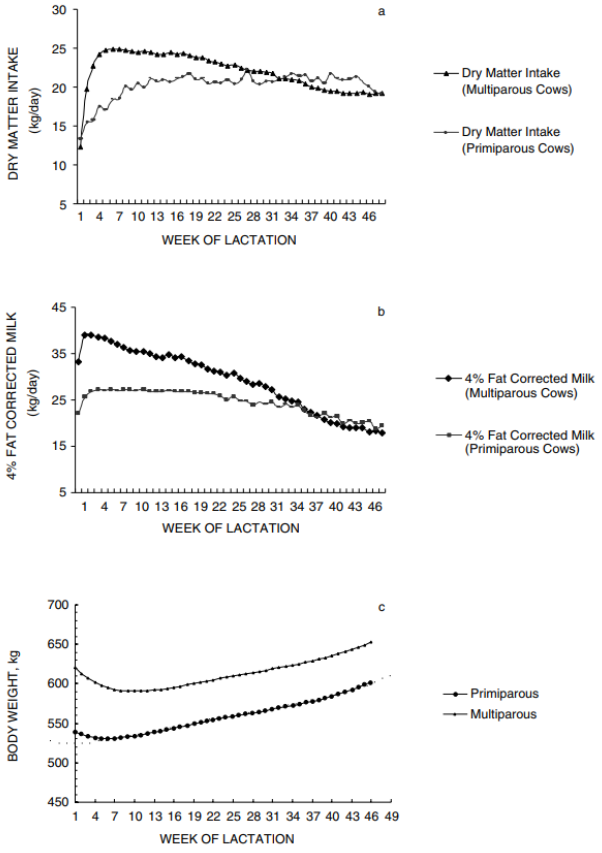


Figure 4 Illustration of changes in dry matter intake, milk yield, and body weight in primiparous as compared to multiparous cows over the course of lactation. Figure from Council (2001)

Many studies have been conducted to improve our understanding of the processes involved in the complexity of this biological system. Five important physiological factors that affect the

variation of FE are: 1) feed intake, 2) digestion of feed, 3) metabolism, 4) physical activity of the animal and 5) thermoregulation (Herd & Arthur, 2009). These physiological mechanisms have been thoroughly discussed and explained by Herd and Arthur (2009) and compared in dairy cows in divergently high and low RFI groups by Xi et al. (2016). Even earlier, in beef cattle, E. Richardson and Herd (2004) summarized many biological/physiological mechanism that contribute to the complexity of RFI trait in beef cattle (Figure 5). Xi et al. (2016) reported that the RFI differences resulted in difference in fat mobilization (body size), metabolites (urea nitrogen and serum level of leptin), milk production performance as well as feeding behaviour.

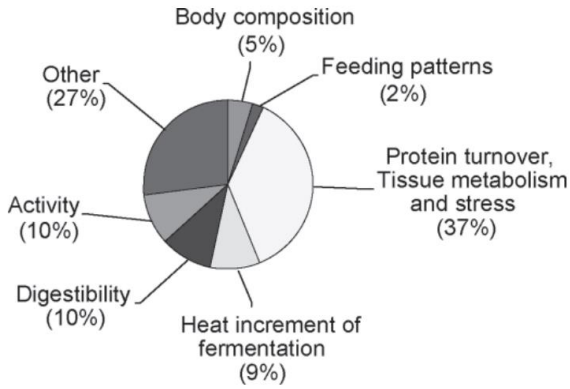


Figure 5 Pie chart showing the physiological mechanisms that contribute to variation in RFI in beef cattle, which were selected according to RFI. Chart from E. Richardson and Herd (2004)

In spite of the difficulties in accurate determination of RFI, a heritability for RFI has been established in dairy cattle with estimations from around 0.01 to 0.38 (Berry & Crowley, 2013; Tempelman et al., 2015; Veerkamp et al., 1995), 0.47 repeatability (E. Connor et al., 2013) and 0.40-0.43 reliability (Pryce et al., 2014). Promising results regarding the use of RFI for breeding purposes have been obtained in beef cattle, where selection of animals according to RFI decreased the feeding costs (Xi et al., 2016). The RFI can be designed to take almost every possible factor that contributes to FE into account, including enteric formation and release of greenhouse gasses with associated negative environmental effects.

The biology underlying FE in dairy cattle is as mentioned very complex, and in fact for no other trait of livestock production do the correlations between major traits used to calculate a breeding index change over time to the extent that they do for FE in dairy cattle (Pech et al., 2014). For that reason, RFI appears to be a better and useful term to include in dairy cattle breeding as a measure of FE. The use of this term in practice is presently complicated, however, due to the limited number of observations on actual FI in dairy cow populations, and hence a low accuracy in the genetic prediction of FI, which must be known to calculate RFI. Therefore, selection

according to RFI, which composite all important factors that contribute to FE, could potentially contribute to improve the accuracy in selection of more feed efficient dairy cows. Thus, in order to understand the mechanism of this complexity, it is very important to investigate further the physiology in relation with gene expression specifically related to RFI irrespectively of the stage of lactation.

Omics technologies and systems biology approach applied on RFI could particularly in dairy cattle breeding be a way forward to overcome some of the challenges involved with genetic prediction of the important traits underlying FI.

Previous omics studies in dairy cattle have, as previously mentioned, focussed on mammary tissue or milk, and the mammary gland is obviously the organ responsible for synthesis and secretion of milk. However, in relation to FE and RFI, the liver is a crucial organ responsible for the coordinated distribution and metabolism of absorbed or mobilised nutrients to other tissues (Drackley et al., 2001). Before any delivery to other tissues in the body, absorbed nutrients (except for lipids) will pass through the liver. The liver has vital coordinating functions in the body such as glucose supply, amino acid and nitrogen excretion, as well as fatty acid oxidation, and immunity (Herd & Arthur, 2009; Zachary & McGavin, 2013). If the cow cannot increase feed intake sufficiently after calving, a resulting extensive fat mobilization in early of lactation can give rise to development of fatty liver and ketosis with negative impact on feed intake, milk production and animal health (Herdt, 2000). The liver is therefore the most widely used tissue used in gene expression studies in relation to feed consumption and FE in other livestock species and beef cattle (Alexandre et al., 2015; Ramayo-Caldas et al., 2018; Zarek et al., 2017). The liver is also responsible for synthesis of all plasma proteins, clotting factors, and enzymes involved in detoxification (Seal & Reynolds, 1993). The so-called acute phase proteins responsible for the first line of defence to infections such as mastitis, are among the plasma proteins synthesized in the liver.

Therefore, application of the omics technologies and new systems biology approaches on liver could potentially provide new important information about genes and coordinated biological mechanisms underlying differences in FE, and specifically the RFI trait.

Why is it relevant to characterize RFI differences across different breeds?

Various dairy breeds are being used worldwide. Among them, the Holstein and Jersey breeds have been widely used and are among the top producers all over the world. In Denmark, Holstein and Jersey are the top dairy breeds (RYK, 2017). Holstein and Jersey cows are similar with respect to feed efficiency background (Blake et al., 1986); however, from the milk producers'

perspective they have several other breed specific characteristics. Generally, Holstein cows produce more milk compared to Jersey, among other things due to a larger body size and feed intake. On the other hands, Jersey cows are known for their excellent reproductive efficiency (Felippe et al., 2017) and robustness.

Many studies have been conducted on these two breeds to characterize their productivity, fertility and digestive function. However, there are no omics or molecular biology studies characterizing similarities and dissimilarities between the two breeds, and no studies have addressed which genetic markers or candidate genes are responsible or linked to differences in FE and RFI between different breeds. Moreover, if information can be retrieved in omics studies regarding new biomarkers, we obviously need to know, whether the same biomarkers apply to different breeds. Therefore, it is interesting to investigate and compare these two breeds further.

Why is it interesting to study RFI changes in relation to changes in diet composition?

The diet or feed ration composition is an important factor determining feed intake and hence FE. This will in turn determine whether nutrient requirements of dairy cows can be fulfilled for different life manifestations such as body maintenance, pregnancy, live weight change and milk production (Council, 2001).

The nutrients supplied by the diet can alter the gene expressions profiles either directly or indirectly, and subsequently affect synthetic processes and metabolic and/or signalling status of different cells, tissues, organs. Thus feed composition can affect physiological functions in the body by interacting or inducing changes in the transcriptome, this concept has revolutionized the field of ruminant nutrition, including dairy cows (Bionaz, 2014). Several studies reported that different types of diet or feed restrictions could affect fertility and milk quality and quantity, by altering the profile of mRNA expression (Bionaz & Loor, 2012; W. Li et al., 2013; Loor et al., 2007; Velez & Donkin, 2005).

An important factor relating to feed composition of dairy cow diets is the impact of the concentrate to forage ratio on dry matter intake (Council, 2001). Manipulation of dietary energy can be a potential way to alter the nutrient availability, metabolism and hence milk production of dairy animals (Dann et al., 2006). As long as forestomach health is not negatively affected, high concentrate or high grain diets will lead to overall increased dietary DM intake due to shorter retention time of the feed in the rumen, and greenhouse gas emission per kg of milk produced or per kg of DMI can be reduced due to changes in the patterns of rumen fermentation.

There are few studies that have compared the effect of different rations on the mRNA expressions profiles in dairy cows. In one study, which tested the effect of a high concentrate

diet on the rumen papillae of Holstein cows, it was found that genes related to inflammation were upregulated (R. Zhang et al., 2016). A study on the hepatic gene expression between Holstein cows fed with total mixed ration and pasture showed differences in *IGFBP6* and *LEPRb* mRNA expression (Astessiano et al., 2015). The abundance of *IGFBP6* mRNA expression in this study could probably be ascribed to undernourishment of the animals (Wilkinson, 2011) and were associated with lower fertility during the transition period (Cummins et al., 2012). This shows that different dietary energy levels might affect gene expression in different tissues, which in turn can have implications for animal performance. However, these studies only included very few specific genes. A study on the whole mRNA expression profile in a tissue would provide much more insight into the genes behaviour and patterns of expression.

In conclusion, throughout the literature from different species, including dairy cattle, several biological pathways have been unravelled that illustrate the complexity of a multifactorial trait such as FE. FE is obviously highly affected by diet, and dietary intake is also highly affected by stage of lactation and linked to milk production, which subsequently makes the evaluation and investigation more complicated. We are still lacking information and studies that can link genetic traits to these issues in dairy cattle. Transcriptomics and new systems biology approaches could help to improve our understanding of the genetic features that are linked to ability to cope with changes in milk production and energy balance over the course of lactation. Therefore, by characterising the genetic networks linked to diet induced changes in gene expression patterns, we could improve the accuracy in genetic prediction of FE, as defined e.g. by RFI. The RNA-seq transcriptomics and systems biology research specifically have never previously been applied to dairy production, and the new knowledge discovered by application of these techniques, could provide new biomarkers and facilitate development of more accurate breeding program for high FE dairy cows in the future.

Materials and methods

In this chapter, the experimental design, animals, dietary treatments, tissue samplings and the bioinformatics pipeline applied to liver samples in this PhD project will be described. The PhD project workflow is presented in Figure 6.

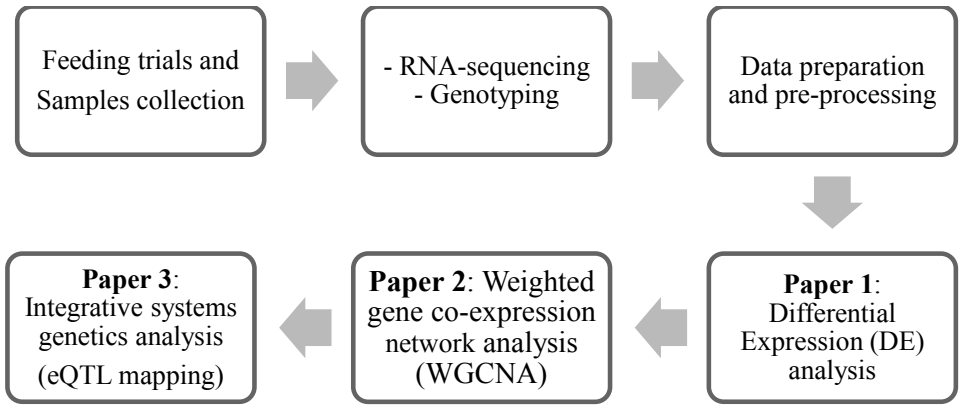


Figure 6 Experimental workflow for the identification of candidate genes for RFI in Nordic dairy cattle

It was possible to obtain biological samples (liver biopsies) from cows used in an animal feeding experiment, designed to improve utilization of feed resources in two dairy breeds. The experimental animals and the experimental design of that experiment have previously been described by B Li et al. (2016) and Salleh et al. (2017). In the following, only a brief description of the animal experiment will be given, and the major emphasis will be on the experimental procedures undertaken as part of this PhD, namely liver biopsy and blood sampling, and the omics and systems biological analyses applied to the samples and the processing of the data from these analyses.

Experimental animals and dietary treatments

The animal feeding experiment was conducted at the Danish Cattle Research Centre (DCRC) in Aarhus University, Foulum, Denmark, where a total of 200 experimental cows were several parameters relating to feed efficiency, including RFI, were evaluated and calculated as previously reported (B Li et al., 2016). For this PhD study, we selected cows based on their RFI, which was calculated based on DMI regressed with weeks of lactation, the management group in which the cows were held, and the interaction between weeks of lactation, breed and parity. Liver samples were obtained from 20 percent samples of the whole population of cows in the experiment, namely 10 percent from the extreme low-RFI cows and 10 percent from high-RFI

cow (Figure 7). The low-RFI cow will also be referred to in the following as high FE cows, and high-RFI cows will be referred to as low FE cows.

Figure 7 Graph of normal distribution of feed efficiency value (RFI value) with the top and bottom 10% extremes. The 10 % extremes were selected for the RNA sequencing to identify candidate genes for biomarker development (DEG, CEG and eQTLs).

Initially a total of 10 Holstein and 10 Jersey were selected among primiparous and multiparous cows based on their RFI value in order to create two groups of cows with divergent RFI for each breed, but one Holstein cow was dropped from the study because the liver biopsy could not be taken. A summary of the final number of cows in each of the 4 treatment groups is presented in Table 2.

Table 2 Number of animals used in the experiment

	HOLSTEIN	JERSEY	Total
Low RFI	5	5	10
High RFI	4	5	19
Total	9	10	19

The two FE groups were exposed to two different dietary treatments over the course of the experiment, namely a Control (C) diet and a High Concentrate (HC) diet with 68:32 and 39:61 forage:concentrate ratio, respectively. The diets were fed to the cows in two different periods so that each cow was exposed to both diets. The cows were each experimental diet for a 14-26 day period, and the liver biopsies were obtained on the last day of each period. The feeding trials and sampling from the cows used for this PhD project were conducted in five separate time blocks including 4 cows. This design was due to respiration chamber measurements in the larger experiment, and limited number of respiration chambers (4). Table 3 presented the allocations of the cows during the feeding trials.

Table 3 Scheme illustrating the allocation of cows on diets in each experimental block

Cow ID	Block	Breed	RFI	Diet	
				Period 1	Period 2
6005	1	Holstein	Low	C	HC
5957	1	Holstein	High	HC	C
6020	1	Jersey	High	C	HC
6004	1	Jersey	Low	HC	C
5729	2	Jersey	High	HC	C
5802	2	Jersey	Low	C	HC
5682	2	Holstein	High	C	HC
5751	2	Holstein	Low	HC	C
6095	3	Jersey	Low	C	HC
6162	3	Jersey	High	HC	C
6144	3	Holstein	Low	HC	C
6118	3	Holstein	High	C	HC
5544	4	Holstein	Low	C	HC
5790	4	Holstein	High	HC	C
5691	4	Jersey	High	C	HC
5739	4	Jersey	Low	HC	C
6090	5	Jersey	High	HC	C
6160	5	Jersey	Low	C	HC
6199	5	Holstein	Low	HC	C
6167[1]	5	Holstein	High	C	HC

Biological sample collection

For this PhD project liver biopsies and blood samples were collected from each animal, while placed in tie-stalls (Picture 1). The samples were collected early in the morning before the morning feeding.

The liver biopsy samples were taken from each cow in each of the two experimental feeding periods and were sequenced for the RNA expression. First the skin surrounding the area of the incision was shaved and cleaned. Approximately 10-20 mg of liver were withdrawn using a biopsy gun after the cows had first been locally anaesthetised using 10 millilitres (ml) of Procamidor®vet (20miligrams (mg/ml)) underneath the skin around the intercostal muscle (Picture 2). Immediately after the withdrawal of the liver tissue from the animals, the liver biopsies were immersed in RNAlater solution and stored at 4°C for approximately one week.

[1] Holstein cow that was dropped from the study because of the liver biopsy failed during the first period

Then, after one week, the RNAlater solution was removed from the biopsies and the samples were stored at -80 °C until further processing.

Picture 1 The experimental cows at the tie-stalls area during the milking period

The blood was withdrawn by venepuncture from a jugular vein and sampled into 10 ml EDTA coated vacutainer tubes. The full blood samples were stored at -20°C until further processing.

Picture 2 Liver biopsy sampling

RNA-sequencing

Briefly, the mRNA was extracted from the liver sample. The liver tissue was first disrupted and homogenized with TissueLyser II, Qiagen, together with 1 ml of Qiazol lysis reagent on a bead mill for three minutes at 30 Hz. Subsequently, the mRNA was extracted following the manufacturers procedures using RNeasy® Mini Kit and MaXtract High Density. The quality and quantity of the RNA were measured using an Agilent 2100 Bioanalyzer and NanoDrop® ND-1000 spectrophotometer, respectively, before further cDNA library preparation. RNA was sequenced by AROS Biotechnology A/S (Denmark) using an Illumina HiSeq 2500 machine.

DNA Genotyping

The full blood samples were genotyped by Neogen GeneSeek® (Lincoln, NE, USA) using 777k BovineHD BeadChip (Illumina, Inc., San Diego, CA, USA).

Datasets preparation and processing

Briefly, the mRNA sequencing raw data and genotype data were processed using a superclusters with Operative system: openSUSE 13.1 Bottle (x86_64), Linux version: 3.11.10-7-desktop, RAM: 504 GB, #CPUs: 64.

The RNA raw reads were pre-processed using FastQC version 0.11.3 (Bioinformatics, 2011). The reads were aligned to the Bovine reference genome release 82 using STAR aligner (Dobin et al., 2013). After the alignment, quality control of the mapped reads was done using Qualimap version 2.0 (Okonechnikov et al., 2015). Then the HTSeq-count tool was used to compute the gene expression counts (Anders et al., 2014).

The genotype data was processed using PLINK 1.90 beta software (Chang et al., 2015; Purcell et al., 2007). The raw genotype data was filtered by Hardy Weinberg Equilibrium (HWE < 0.0001), Minor Allele Frequency (MAF< 0.15), and missing genotype rates (mind > 0.1). The genotype data were also pruned to remove SNPs in strong linkage disequilibrium. A total of 536,420 SNPs was removed after the filtration procedure. The remaining 241,542 SNPs were used for the rest of the analysis. Subsequently, the pre-processed datasets were used in the eQTL mapping analysis.

All other subsequent analysis was performed using R studio software version 3.3.1.

Differential gene expression analysis[2]

The identification of differentially expressed genes (DEGs) was performed by using DESeq2 package version 1.12.0 (Love et al., 2014). Firstly, low gene counts were filtered out by removing genes with less than one count per million (cpm) in 90 percent of the sample size. The

[2] The details of the procedure can be found in Paper 1

generalized linear model was fitted in DESeq2 including potential confounding effects: diet and parity number. The DE analysis was performed separately for the two breeds.
These two datasets were analysed by using two types of models:

Model 1: with interaction effect

$$Y = Parity\ number\ +\ Diet\ +\ RFI\ +\ Diet{:}\,RFI$$

Model 2: without interaction effect

$$Y = Parity\ number\ +\ Diet\ +\ RFI$$

The gene were considered DE when the False Discovery Rate (FDR) P-value was lower than 0.05.

Weighted gene co-expression network analysis (WGCNA) [3]

Gene co-expression network analysis were done by using the WGCNA package (Langfelder & Horvath, 2008). The two breeds were analysed separately.
The gene expression counts were used to generate an unsigned network. Briefly, the adjacency matrix was constructed by computing the pairwise Pearson Correlation Coefficients (PCCs). Next, a topological overlap matrix (TOM) was used as an interconnectedness measure to define the modules. Modules of highly connected genes in the network were identified by a dynamic tree cut algorithm. Each module was arbitrarily labelled with a colour. The module trait relationship (MTR) was computed as the correlation between the first principal component of the gene expression values in each module (Module Eigengene) and the trait (RFI, Diet and Parity number). The MTRs were evaluated to select significant (FDR p-value<0.05) associations of the modules and traits of interest.
In addition, for each significant module, hub genes were identified by selecting the high value of Module Membership (MM>0.8).

[3] The details of the procedure can be found in Paper 2

Functional enrichment analysis

Functional enrichment analysis was performed for both DEG and CEG independently. This analysis was also conducted separately for each breed.

Differentially expressed genes (DEG)

The list of DE genes that was assigned to functional enrichment analysis using the GOSEQ package in R to account the length bias (Young et al., 2010). No length bias was observed in the set of DE genes. Hence, we further analysed using the gene set enrichment analysis (GSEA) (Subramanian et al., 2005), STRING database (Szklarczyk et al., 2014) and QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, http://www.qiagen.com/ingenuity) to find and identify any significant GO and pathways related to RFI trait. The significant GO and KEGG pathways identified were set at $P<0.05$.

Co-expressed genes (CEG)

The set of genes included in the significant modules for both breeds were also assigned to the functional enrichment analysis. A Cytoscape plug-in, ClueGO v2.2.6 application (Bindea et al., 2009) was used to identify significant GO terms and KEGG pathways. Significant GO terms and KEGG pathways were considered as significant when $P<0.05$ using Benjamini Hochberg multiple testing correction. The upstream regulators were identified by using Ingenuity® Pathway Analysis (IPA®).

Integrative genomics analysis (eQTL mapping)

We integrated transcriptomics and genomics data to identify genetic variants associated with the expression of candidate genes (DEGs and HGs) for RFI. Genetic variants that are associated with gene expression are called expression QTL (eQTL). This was done by integrating the genomic data obtained from genotyping together with candidate genes from the RNA-seq analysis.

To identify possible significant variants for the RFI trait, eQTL mapping was performed for a total of 170 the candidate genes (DE genes and hub genes) by using the R package MatrixEQTL v 2.1.1 (Shabalin, 2012). The MatrixEQTL package performed the analyses separately for local eQTLs (SNP within 1Mb from the targeted genes) and for distant eQTLs (different chromosome). Firstly, the RNA-seq data were merged together between the Holstein and Jersey cows by only keeping genes they had in common and filtering out low count genes. A total of 160 genes survived after this filtration procedure. The breed and the parity number were included

as covariate in the analysis. Subsequently, following the default procedure recommended by the developer, and log2 transformation of the gene count was used for the eQTL mapping.

Comparison of the eQTL with the Animal Genome cattleQTLdb

The eQTLS that were significantly associated with the expressions of the candidate genes were further analysed and compared with previously ereported QTL related studies. This analysis was done by comparing the results with the data from the Animal Genome cattleQTLdb database (Hu et al., 2015). The cattle QTL and association data (UMD_3.1 in GFF3 format) was downloaded from the AnimalQTLdb website (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index). Afterwards, long QTL regions and SNPs that had more than one flankmarker were filtered out from the data. In total, 94322 SNPs were used for the comparison of the eQTLs. The SNPs information was from 337 studies, 63 breeds, and 366 traits of 6 trait types.

The comparison was done by mapping the identified significant local-eQTL in the cattleQTLdb. A flanking region of 500 Kb included around the SNP position. The QTLs overlapping for at least one nucleotide were considered as a match.

General Results

This chapter presents the general results of the RNA-seq and genomic data analysis, which consisted of the Differential Gene Expression (DGE) analysis (Paper 1), Weighted Gene Co-expression Network Analysis (WGCNA) (Paper 2) and the integrative genomics analysis (Paper 3).

Before we further analysed the gene expression profiles for the datasets from both breeds, we plotted the Principal Component (PCA) plots in order to observe and understand the variation of the samples by classifying the samples according to their FE groups (high- and low-RFI corresponding to low- and high-FE, respectively) and Diet groups (C and HC).

The PCA plots in Figures 8 and 9 represent the main variations captured by the first and second component using the entire expression profile of the samples. Samples are indicated with different shapes and colours to distinguish between diets and FE groups, respectively.

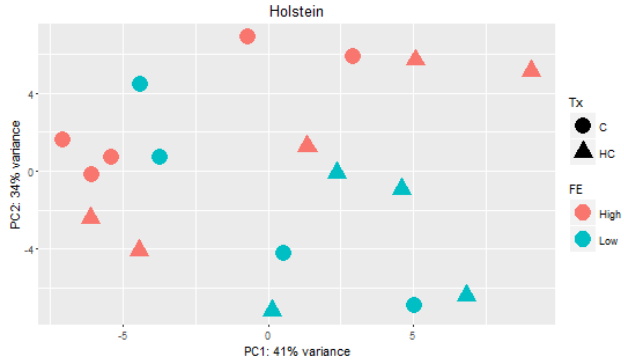


Figure 8 PCA plot for the Holstein breed; samples classified by diet (Tx) and feed efficiency (FE). C-high FE = red circle; HC-high FE= red triangle; C-low FE= blue circle; HC-low FE= blue triangle

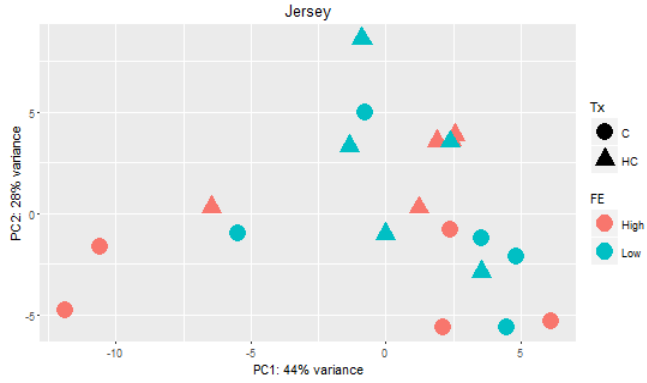


Figure 9 PCA plot for the Jersey breed; samples classified by diet (Tx) and feed efficiency (FE). C-high FE = red circle; HC-high FE= red triangle; C-low FE= blue circle; HC-low FE= blue triangle

Paper 1: Differential Gene Expression (DGE)

Subsequently, the DGE analysis was performed to identify differences in the gene expression profiles between high- and low-RFI group and to test the interaction effect between RFI and Diet (RFI x Diet) on the gene expression. We identified 70 genes (30 downregulated and 40 upregulated in the high-RFI group) and 19 genes (10 downregulated and 9 upregulated genes in the high-RFI group) that were differentially expressed in Holstein and Jersey cows, respectively. Two genes (*SEC24D* and *FIZ1*) were significantly affected by the interaction between diet and RFI groups in Jersey cows. The list of the DEGs is presented in Tables 4 and 5.

Table 4 Differentially expressed genes list in Holstein cows

	Ensembl.Gene.ID	Associated.Gene.Name	baseMean	log2FoldChange	padj
1	ENSBTAG00000000170	Uncharacterized protein	226.93	-0.43	1.34E-02
2	ENSBTAG00000000654	*ARMC4*	58.66	-0.59	3.58E-04
3	ENSBTAG00000001009	*HCLS1*	440.58	0.32	2.80E-02
4	ENSBTAG00000001154	*DGAT2*	511.45	-0.37	2.92E-02
5	ENSBTAG00000001204	*KIAA1462*	225.21	-0.42	1.76E-02
6	ENSBTAG00000002224	*UHRF1*	77.59	-0.50	2.09E-04
7	ENSBTAG00000002526	*BDH2*	1382.09	-0.58	3.67E-16
8	ENSBTAG00000002705	*REC8*	304.68	-0.37	1.71E-03
9	ENSBTAG00000003696	*CCDC64*	45.09	0.45	2.70E-02
10	ENSBTAG00000003718	*HACL1*	6329.55	0.32	3.87E-02
11	ENSBTAG00000004076	*OXER1*	223.68	-0.44	1.05E-02
12	ENSBTAG00000004558	*C15orf48*	89.06	0.51	3.37E-03
13	ENSBTAG00000004908	*CHRNE*	246.50	-0.75	6.38E-08
14	ENSBTAG00000005287	*CYP7A1*	4126.21	0.46	1.31E-02
15	ENSBTAG00000005629	*AIM1L*	913.59	-0.30	3.37E-03
16	ENSBTAG00000006452	*CD3D*	77.75	0.41	3.66E-02
17	ENSBTAG00000006599	*INHBE*	605.95	-0.424	4.41E-02
18	ENSBTAG00000006675	*PCSK6*	3039.35	-0.19	2.80E-02
19	ENSBTAG00000006934	*CYP11A1*	649.88	0.49	4.84E-03
20	ENSBTAG00000006978	*HSD17B4*	13797.37	0.30	2.70E-02
21	ENSBTAG00000006999	*RYR1*	148.34	0.52	1.99E-03
22	ENSBTAG00000007554	*IFI6*	136.81	0.38	4.90E-02
23	ENSBTAG00000007828	*SLA*	118.26	0.32	3.87E-02
24	ENSBTAG00000007895	*SLC20A1*	880.08	-0.56	1.36E-05
25	ENSBTAG00000008160	*MBOAT2*	440.60	0.34	4.39E-02
26	ENSBTAG00000008424	*ABR*	459.60	0.33	3.18E-02
27	ENSBTAG00000008913	*TMEM98*	333.14	-0.52	1.00E-03
28	ENSBTAG00000009085	*SLC35A5*	1691.36	0.28	6.65E-03
29	ENSBTAG00000009137	*NKG7*	215.44	0.38	2.80E-02
30	ENSBTAG00000009263	*MFSD1*	2661.74	0.24	1.31E-02
31	ENSBTAG00000010463	Uncharacterized protein	394.14	0.38	2.46E-03

32	ENSBTAG00000010564	*ELOVL6*	994.74	0.44	2.70E-02
33	ENSBTAG00000011771	*FICD*	107.91	-0.36	3.72E-02
34	ENSBTAG00000011832	*ALDH18A1*	404.91	0.31	2.60E-02
35	ENSBTAG00000012007	*SOCS2*	835.32	0.42	4.39E-02
36	ENSBTAG00000012995	*CCDC109B*	52.93	0.42	3.72E-02
37	ENSBTAG00000013596	*NR1H4*	1215.49	0.24	1.20E-02
38	ENSBTAG00000014064	*FGFR2*	1554.53	-0.40	1.14E-03
39	ENSBTAG00000014791	*CTH*	224.75	-0.54	1.74E-06
40	ENSBTAG00000015313	*CEACAM19*	51.60	-0.94	1.81E-14
41	ENSBTAG00000015419	*ARHGEF37*	204.98	0.47	1.99E-03
42	ENSBTAG00000016542	*LAMB3*	1783.59	0.43	2.47E-02
43	ENSBTAG00000017567	*ACACA*	844.77	0.40	5.07E-03
44	ENSBTAG00000018116	*MTFP1*	88.16	-0.39	2.60E-02
45	ENSBTAG00000018548	*INTS7*	6522.27	0.24	1.99E-03
46	ENSBTAG00000018604	*SEMA4G*	4847.00	-0.16	4.90E-02
47	ENSBTAG00000018723	*SLC25A34*	96.51	-0.44	2.92E-02
48	ENSBTAG00000019585	*MYOM1*	962.39	0.45	2.51E-02
49	ENSBTAG00000020116	*JSP.1*	2041.87	0.32	1.68E-02
50	ENSBTAG00000020371	*ACOT8*	312.40	0.42	1.34E-02
51	ENSBTAG00000020375	*CYP2C9*	5129.18	0.42	3.11E-02
52	ENSBTAG00000020499	Uncharacterized protein	68.21	0.57	3.58E-04
53	ENSBTAG00000020755	*SELP*	478.82	-0.39	1.50E-02
54	ENSBTAG00000021746	*ANXA5*	333.86	-0.38	3.87E-02
55	ENSBTAG00000023851	*FAM102A*	229.01	-0.51	1.42E-03
56	ENSBTAG00000023929	*FOSL2*	189.62	0.42	2.80E-02
57	ENSBTAG00000024044	*CDKL4*	82.12	0.52	1.99E-03
58	ENSBTAG00000025258	Uncharacterized protein	102.69	0.54	1.71E-03
59	ENSBTAG00000025898	*TBC1D8*	442.30	0.27	4.76E-02
60	ENSBTAG00000026779	*LYZ*	516.84	0.64	6.78E-06
61	ENSBTAG00000030966	*TAF6*	419.59	-0.26	1.05E-02
62	ENSBTAG00000035998	*CKB*	332.04	0.39	4.90E-02
63	ENSBTAG00000037913	Uncharacterized protein	436.53	0.21	4.29E-02
64	ENSBTAG00000037917	*SLC17A1*	2786.41	0.44	1.71E-02
65	ENSBTAG00000038496	*CR2*	1355.76	-0.54	3.72E-06
66	ENSBTAG00000038962	*SLC6A11*	2637.35	-0.37	1.00E-02
67	ENSBTAG00000039731	*RND3*	1761.21	-0.25	2.80E-02
68	ENSBTAG00000046076	Uncharacterized protein	124.75	-0.42	4.85E-02
69	ENSBTAG00000046730	Uncharacterized protein	139.83	0.37	4.90E-02
70	ENSBTAG00000047529	Uncharacterized protein	110.84	-0.54	1.90E-03

+v e log2 fold change = upregulated in high-RFI cows
-ve log2 fold change = downregulated in high-RFI cows
padj = Adjusted p-value

Table 5 Differentially expressed genes list in Jersey cows

	Ensembl.Gene.ID	Associated.Gene.Name	baseMean	log2FoldChange	padj
1	ENSBTAG00000006525	*FDXR*	125.97	-0.65	6.21E-13
2	ENSBTAG00000008066	*PKDREJ*	76.75	0.56	1.15E-05
3	ENSBTAG00000013689	*MCTP2*	148.26	0.53	1.07E-05
4	ENSBTAG00000027727	Uncharacterized protein	284.20	0.48	3.73E-04
5	ENSBTAG00000038487	*ZNF613*	155.40	-0.39	2.63E-02
6	ENSBTAG00000046257	*GIMAP4*	650.18	-0.39	2.40E-03
7	ENSBTAG00000005182	*BOLA-A*	434.95	-0.39	1.33E-03
8	ENSBTAG00000014402	*GIMAP8*	713.05	-0.38	8.57E-03
9	ENSBTAG00000045727	Uncharacterized protein	921.10	0.38	3.25E-02
10	ENSBTAG00000019026	*EXTL2*	34.59	0.38	3.72E-02
11	ENSBTAG00000037440	*ZNF197*	281.10	0.36	1.64E-02
12	ENSBTAG00000021751	*RASEF*	36.02	-0.35	1.06E-02
13	ENSBTAG00000027205	*PGBD5*	30.06	-0.34	2.63E-02
14	ENSBTAG00000031737	*TMEM102*	26.53	0.34	3.72E-02
15	ENSBTAG00000009087	*GNG10*	1516.44	-0.33	2.63E-02
16	ENSBTAG00000040323	Uncharacterized protein	1003.60	-0.32	2.63E-02
17	ENSBTAG00000014161	*ARMC10*	258.84	-0.30	2.63E-02
18	ENSBTAG00000013106	*C19orf81*	26.34	0.30	2.63E-02
19	ENSBTAG00000047379	*CYP3A4*	2422.37	0.29	4.34E-02

+ve log2 fold change = upregulated in high RFI group
-ve log2 fold change = downregulated in high RFI group
padj = Adjusted p-value

It was not possible to find any significant GO terms and KEGG pathways using the GOseq package, because of limited annotations for Bovine. Therefore, the functional analysis results were further interpreted from the STRING, IPA® and GSEA analyses.

Generally, in Holstein cows, pathways that we discovered to be significantly enriched in the DEGs list were similar in the three analysis using STRING, IPA® and GSEA. The top pathways identified with the functional enrichment analysis of the DEGs list using the STRING database were: primary immunodeficiency, Natural killer cell mediated cytotoxicity, T cell receptor signaling pathway, Leukocyte transendothelial migration, Chemokine signaling pathway, FC gamma R mediated phagocytosis as well as Propanoate metabolism.

For Jersey cows, no significantly enriched pathway from the 19 DEG could be identified using STRING, but the GSEA analyses revealed significantly enriched (FDR<0.01) pathways related to leukocyte transendothelial migration, primary immunodeficiency, retinol metabolism, metabolism of xenobiotics by cytochrome P450 and ether lipid metabolism.

Paper 2: Weighted gene co-expression network analysis (WGCNA)

To understand better the behaviour of change of the gene expression profiles, we conducted the WGCNA analysis in order to identify co-expressed genes (CEGs) and hub genes (HGs) that were significantly associated to RFI.

From the separate analysis on both breeds, we identified 13 modules that were significantly associated with the RFI trait in Holstein and three modules in Jersey cows. We tested all the significant modules for functional enrichment using ClueGO application and STRING database. We compared the findings from both tools, and observed that the results and outputs from both platforms were similar.

Among all the significant modules associated with RFI and Diet, we found that only the top modules were biologically meaningful (significant biological enrichment). The top significant module was the salmon module (MTR RFI= 0.7) in Holstein with a positive association to the RFI trait, and the lightsteelblue1 module (MTR RFI= -0.57) in Jersey with a negative association to the RFI trait. The top significant module associated with different diet (forage:concentrate) was Magenta module (MTR Diet=0.82) for Holstein and Violet module (MTR Diet=-0.47) for Jersey.

The gene list from the salmon module in the Holstein cows was enriched for biological processes relating primarily to lipid metabolism (i.e. cholesterol biosynthetic process, steroid biosynthetic process, lipid biosynthetic process, small molecule biosynthetic process, lipid metabolic process and isoprenoid biosynthetic process).

While the gene list in the lightsteelblue1 module in Jersey cows was enriched for biological process relating primarily to immunocompetence functions (i.e. positive regulation of T cell activation, positive regulation of immune system process, chemokine-mediated signalling pathway, positive regulation of interleukin-4 production, positive regulation of T cell proliferation, thymic T cell selection, immune system process, single organismal cell-cell adhesion, T cell co-stimulation).

For module associated with the diet (forage:concentrate), only Magenta module has enriched to the biological meaningful pathway which was related to Triglyceride homeostasis process. On the other hand, for Jersey, the Violet module has limited output from the enrichment analysis. Therefore, only the triglyceride homeostasis process related to the Magenta module will be further discussed in the discussion section.

In the salmon and lightsteelblue1 modules, 52 and 29 genes could be classified as hub genes (HGs) for Holstein and Jersey cows, respectively, with MM >0.80 (Figures 10 and 11).

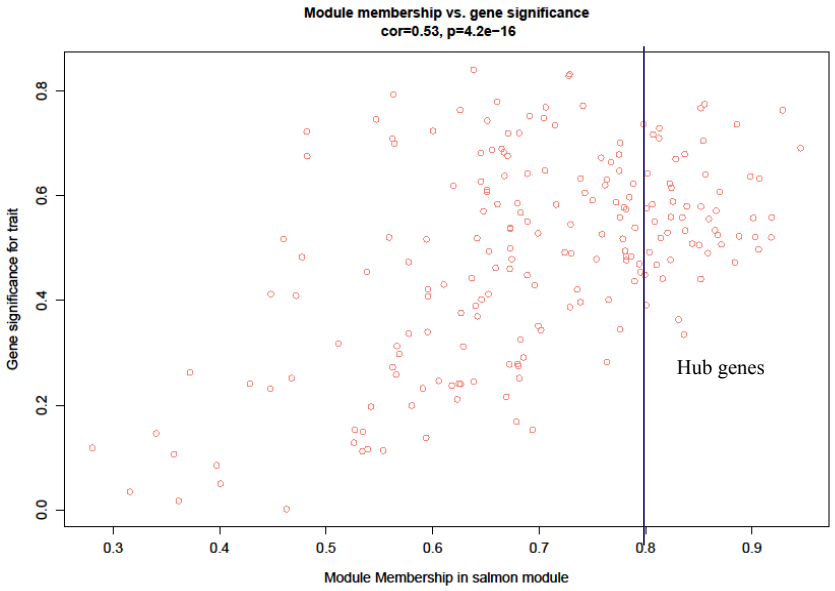


Figure 10 Scatterplot trend of each gene in the Salmon module. The plot shows a positive correlation (r=0.53) between the module membership (x-axis) against gene significance (y-axis).

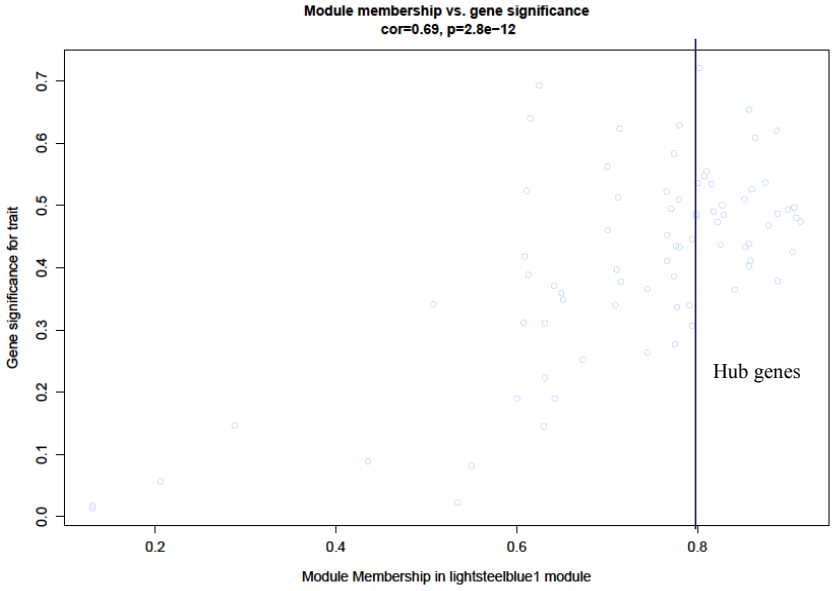


Figure 11 Scatterplot trend of each gene in the Lightsteelblue1 module. The plot shows a positive correlation (r=0.69) between the module membership (x-axis) against gene significance (y-axis).

The lists of the co-expressed genes in the interest modules (salmon and lightsteelblue1) were assigned to the IPA to detect the upstream regulators. We identified several upstream regulators for RFI. In the top significant module in Holstein (salmon module), ATP7B was predicted as

activated, while POR (Cytochrome P450 Oxidoreductase) and cholesterol were predicted as inhibited. In Jersey, the IPA® predicted IFNG (Interferon Gamma) as inhibited and IL10RA (Interleukin 10 Receptor Subunit Alpha), NKX2-3 (NK2 Homeobox 3) and dexamethasone were predicted as activated upstream regulators for the top significant module (lightsteelblue1 module).

Paper 3: Integrative genomics analysis / eQTL mapping

After the identification of candidate genes through DEG and CEG analysis, we further mapped the identified candidate genes onto the genotype data, also called eQTL mapping. This analysis was conducted to identify the eQTL regions that significantly associated to the targeting genes (candidate genes for RFI: DEGs and HGs).

From the eQTL analysis, it appeared that the significant SNPs that associated to the candidate genes in control (low concentrate diet) and high concentrate diet datasets groups were somewhat similar. We identified 20 cis-eQTLs that were significantly associated (FDR<0.05) with seven candidate genes (*BDH2, CHRNE and ELOVL6* for Holstein; *FDXR, CXCL9, CD52* and *GIMAP4* for Jersey). However, there was no significant distant-eQTL (trans-eQTLs) associated with the candidate genes in the analysis performed, when cows were fed the control diet.

On the other hand, when cows were fed the high concentrate diet, we identified 16 local eQTLs SNPs (FDR<0.05) associated with the expression of five genes (*UHRF1, BDH2, HSD17B4,* and ENSBTAG00000047529 for Holstein; only *GIMAP4* for Jersey) and 2891 distant-eQTLs associated with the expression of 45 genes (Supplementary material for Paper 3). Among the local-eQTL, genes that were in common in both diet groups were the *BDH2* and *GIMAP4* genes.

Comparison with previous studies compiled in the CattleQTLdb, revealed that several of the eQTL genes were in the same QTL loci as other traits associated with FE (e.g. DMI, length of productive life, RFI and net merit).

General Discussion

The main focus of this PhD project was to evaluate the utility of omics and new systems biology technologies to identify candidate genes and improve our understanding of the underlying biology of complex multifactorial traits, such as FE in dairy cattle.

To meet the objectives of the study, we characterized the transcriptome of liver biopsies from dairy cows of two different breeds and exposed to different diets and with widely different RFI, and subsequently conducted two types of analyses on this data.

Firstly, DEG analysis was performed to identify differentially expressed genes. Secondly, we applied a system biology approach named co-expression network analysis to identify groups of co-expressed genes associated to RFI. Finally, we integrated the transcriptomics and genomics data to identify expression quantitative trait loci (eQTL) targeting candidate genes for RFI. This chapter will provide a general discussion of the findings from the differential expression analysis, co-expression analysis as well as the eQTL mapping.

Samples and datasets description: Benefits and limitations

In order, to identify candidate genes and obtain a better understanding of the underlying biology associated with RFI and hence feed efficiency, we analysed the gene expression pattern in liver tissue samples. The liver is the organ responsible for central metabolic coordination, regulating nutrient supply to peripheral tissues for maintenance and productive functions, such as milk synthesis and muscle or fat depositions (Seal & Reynolds, 1993). The liver is a delicate tissue in terms of analyses of RNA, because it is rich in nucleases that can degrade the RNA faster than most other tissues with less RNAse activity like muscle and brain (Bauer, 2007; Sidova et al., 2015). Therefore, the RNA degradation needs to be addressed carefully. We used RNA later to reduce post-sampling RNA degradation. By this procedure we ensured that RNA integrity of high quality (RIN value>8).

We made PCA plots to investigate the main variation in the gene expression profiles. The PCA plots could not establish any clear separation between the two RFI classes in none of the breeds, or a clear separation according to diet fed to cows, whereas the cows grouped quite well in relation to parity number (not shown). The parity number status of the cows was from first to third lactation. The different parity was one of the limitations in our study, because experimental groups were not homogenous. However, we have carefully taken all possible biases in consideration that could influence the findings in our analysis. The first thing we made sure would not affect our results, was therefore the parity of cows. Hence, it was important to include

the parity number in all the model in order to remove any potential confounding effects, when testing for DEGs in relation to breed and diet to obtain as precise and accurate results as possible.

This animal experiment was conducted at a research farm at Aarhus University, Denmark; hence, most other environmental factors such as feeding, temperature, milking time etc. were well-controlled. The research farm facilities, such as the barn, tie-stalls, respiration chambers, milking machines and experienced technical staffs allow us to measure and collect all the samples according to the original plan for the experiment. This is very important to make sure the research was conducted according to the rules and regulation as well as avoid or minimize possible bias that can influence the research findings.

The use of the RNA-seq procedure has been widely used recently. Several procedures have been developed and established to allow us to identify candidate genes using the new experimental methods and tools available. This study is a novel approach to use RNA-seq transcriptomic analysis to identify candidate genes for RFI and hence FE in Danish dairy cattle, specifically. Generally, the breeds used in this study are breeds that are commonly used in dairy production worldwide. These two breeds play an important role in milk product industries in Denmark specifically. However, there are only few studies available, which have focused on the identification of candidate genes in dairy animals for feed efficiency, especially in the Jersey breed.

Although the animal number (sample size) in the PhD experiment was relatively small (five samples in each FE group), it has in this study been possible to identify potential new biomarkers and breed specific differences relating to a complex multifactorial trait, such as feed efficiency in dairy cows. According to recommendations from Schurch et al. (2016); Seyednasrollah et al. (2015), a minimum number of five samples are needed to identify DEGs in the DESeq2 package, which was used to identify the DEG in this study. The minimum sample size for co-expression networks analysis from RNA-seq data has been reported to be 20 samples (Ballouz et al., 2015; van Dam et al., 2017), which corresponds to the number of samples in Jersey (20), although not quite in Holstein (18) in the present study. In the data analyses, corrections were made for every possible confounding effect that might have influenced the findings, and the results of the present study do represent an important new contribution in the field, since very few studies have compared the effect of different rations on mRNA expressions profiles in any tissue in dairy cows. In one study, which tested the effect of a high concentrate diet on the rumen papillae of Holstein cows, it was found that genes related to inflammation were upregulated (R. Zhang et al., 2016).

However, before further usage of any information from this study, the findings obviously need to be confirmed and validated in larger population of cows as well as using other types of test. This is to make sure that the candidate biomarkers or biological pathways related information is applicable to the actual application in the future.

RFI was used as a measure of FE in this study, and it is quite well-established and has been used for breeding programs in other types of livestock, especially pig, chicken and beef cattle. In dairy cattle, there is a limitation availability of studies in relation to genetic foundation for differences in feed efficiency, and not least RFI. However, information from other species as well as human orthologous information can be used in to investigate and understand the physiological significance of the findings made in this study for dairy cattle.

Differentially expressed genes (DEGs)

The DGE analysis is the most common approach to identify possible causative changes or differences in gene expression in different cell/tissue populations and/or experimental conditions (Dündar et al., 2015). Prior to other complex analysis, basic analysis should be applied in order to understand the fundamental status of the traits or conditions of interest. The DGE analysis can provide an understanding of the gene expression differences relating to FE in dairy cattle and a basis for elucidating the mechanisms of action and biological functions of the DEGs. A specific hypothesis for the study of the liver was that gene expression profiles in the liver differ between high- and low-RFI animals and that the main DEGs are involved in and responsible for biological mechanisms that are of primary importance in regulating RFI.

Among the list of significantly DEGs, the top significant DEGs that were associated to RFI in Holstein cows were *ACACA, DGAT2, BDH2, ELOVL6* and *CYP11A1*, while in Jersey cows the DEGs were *GIMAP4, GIMAP8, CYP3A4* and *FDXR*. The functions of these top significant genes and their possible relation to the FE trait will be further discussed.

The *ACACA* gene in Holstein was found to have a central place in the STRING analyses, since it was connected to more DEG (ten) than any of the other DEG. These connections included the four other top significant DEGs, which are all part of the *Metabolic pathways* in the functional enrichment analyses. The *ACACA* gene is highly enriched in adipose and mammary tissue, where the enzyme is responsible for *de novo* synthesis of fatty acids (Jensen et al., 1991; Weber et al., 2016; H. Zhang et al., 2015). In our study, the *ACACA* gene was found to be upregulated in the liver in high-RFI (low FE) Holstein cows. Interpretation of this finding in relation to FE of Holstein cows is complicated by the fact that gene expression profiles can be very tissue specific. Thus, in a sheep study by Ropka-Molik et al. (2017), gene expression patterns of lipogenic

genes, including *ACACA*, changed in a differential manner in liver and two adipose tissues in response to changes in the diet. Nevertheless, Weber et al. (2016) reported that a network of lipogenic genes, including the *ACACA* gene, was upregulated in adipose tissue of high- as compared to low-RFI Angus sires, similar to the findings in liver in the present project, and the authors suggested this could be related to a reduction of body fat content in the progeny of low-RFI (high FE) sires. If this is also the case for Holstein cows, then upregulation of the *ACACA* gene in the liver could indicate that high-RFI (low FE) cows may have a higher propensity for directing nutrients into fat deposition rather than milk production. The *ACACA* gene is also a key regulator in *de novo* fatty acid and hence lipid synthesis in the mammary gland, however, the rate of lipid synthesis in adipose and mammary tissues are normally inversely regulated during the course of lactation Baumgard et al. (2017). It would therefore be very interesting to assess gene expression patterns also in mammary tissue to establish, whether increased *ACACA* expression in liver (and adipose) tissues could be associated with decreased expression in the mammary gland in high-RFI cows, and thereby reflect an altered priority in partitioning of nutrients between milk synthesis and body fat deposition.

The *DGAT2* and *ELOVL6* genes were mutually interconnected in the STRING analyses and both upregulated in high-RFI Holstein cows alongside *ACACA*. The *DGAT2* gene encodes for Diacylglycerol O-Acyltransferase 2, which is one of acyltransferases catalyzing the terminal step in triacylglycerol synthesis, and is considered an adipose tissue specific gene with almost undetectable expression in mammary tissue (Bionaz & Loor, 2008). In agreement with this, it has previously been reported that *DGAT2* could be a good candidate gene for meat quality and quantity in pigs (Zang et al., 2016) and ruminants (Fang et al., 2012; J. Li et al., 2009), and it is associated with obesity in humans (H. C. Chen & Farese, 2000). The *ELOVL6* gene encodes for microsomal enzyme that regulates the elongation of C12-16 saturated and monounsaturated fatty acids (Matsuzaka et al., 2012). Expression of *ELOVL6* has been detected in both liver, adipose tissues and mammary epithelial cells in ruminants (S. Chen et al., 2017). Although, de novo synthesis of long-chained fatty acids in ruminants occur predominantly in adipose tissues and not the mammary gland (Tan et al., 2015), upregulation of this gene in mammary epithelial cells *in vitro* was associated with increased synthesis of triglycerids and a shift from C16 to C18 fatty acids in milk fat (Shi et al., 2017). *ELOVL6* has therefore been suggested to have a protective role against lipotoxicíty in the liver and mammary gland induced by C16 (Matsuzaka et al., 2012; Shi et al., 2017), which is the dominating fatty acid mobilized from adipose tissues during negative energy balance.

The *BDH2* gene encodes for the enzyme 3-Hydroxybutyrate Dehydrogenase 2, which is involved in the last step in formation of the ketone body 3-Hydroxybutyrate from acetoacetate, and the first step in metabolism in the reverse proces. This ketone body is partly derived in part from oxidation of butyrate formed during rumen fermentation and partly from synthesis in the liver during periods of negative energy balance. Increased blood levels of ketone bodies in ruminants are normally associated to a disturbance of energy metabolism (Baird, 1982; Bergman, 1971; Gleeson et al., 2016). Downregulation of the *BDH2* in high-RFI (low FE) Holstein cows could potentially reflect a reduced ability for hepatic synthesis of ketone bodies and a reduced ability to oxidize ketone bodies in extra-hepatic tissues. The implications for the cow could in both cases be a reduced ability to cope with extensive body fat mobiliation due to reduced tolerance towards lipotoxic effects of fatty acids in the liver and ketone bodies accumation in systemic circulation.

The final DEG to be discussed is the *CYP11A1* gene, which was not connected to any of the previously mentioned DEGs. The other cytochrome P450s (*CYP*) family genes that were DE in this study were similarly upreguled in high-RFI Holstein (*CYP7A1* and *CYP2C9*) as well as Jersey (*CYP3A4*) cows. The *CYP11A1* gene encodes for the enzyme cytochrome P450scc, which is a member of a superfamily of cytochrome P450 enzymes. P450scc catalyses the conversion of cholesterol to pregnenolone, which is the first step in the tissue specific synthesis of all steroid hormones (glucocorticoids, sex hormones and mineralocorticoids). However, it is not the activity of this enzyme, but rather the cellular supply of cholesterol that is believed to be the rate-limiting step in the formation of these hormones. It can also catalyze cleavage and hydroxylation of the vitamin D3 precurser, 7-dehydrocholesterol, and plant derived ergosterol (Tuckey et al., 2011), which are proceses occurring in the liver and kidneys, but the substrate affinity for the vitamin D precursors are lower than for cholesterol. Without further studies, it is impossible to evaluate whether the upregulation of the *CYP11A1* in high-RFI (low FE) Holstein could point to importance of cholesterol metabolism regulation in the liver, formation of catabolic glucocorticoids in the kidney, sex hormone synthesis or vitamin D-calcium regulation effects associated with RFI and FE.

The list of DEGs for Holstein cows were mostly associated with lipid and ketone body regulatory functions and cholesterol metabolism. All of these functions are also important in relation to regulation of milk fat or membrane synthesis (which is essential for secretion of milk components) in the mammary gland and for partitioning of nutrient between mammary and non-mammary (adipose) tissues. Probably, upregulation of genes in the mammary gland may be associated with downregulation of in adipose tissues to favour milk synthesis rather than deposition in adipose tissues. However, further studies are needed to establish the tissues specific

patterns of gene expression, to be able to establish for sure what these changes imply for other important tissues such as the mammary and adipose tissues.

Nevertheless, this could be definitely an important point to compare gene expression between the different tissues (e.g. mammary gland and adipose tissues) whether the same association will be observed or not in relation with FE trait.

In Jersey cows, only 19 genes were differentially expressed. The *FDXR* gene was the top DEG, and connected to two of the other DEGs, namely GIMAP4 and GIMAP8. All of these 3 DEGs were downregulated in the high-RFI group. *FDXR* or Ferredoxin reductase encodes for a mitochondrial flavoprotein that initiates the first electron transport reaction in all cytochrome P450 catalyzed reactions, receiving electrons from NADPH. This function is important for the generation of free radicals in immune cells. The *FDXR* gene is also involved in other cytochrome P450 enzyme reactions, such as cholesterol metabolic process, oxidation-reduction, steroid and ubiquinone biosynthetic processes (https://www.ncbi.nlm.nih.gov/gene/282604). Interestingly, *GIMAP4* and *GIMAP8* genes have never previously been discussed in relation with feed efficiency. These genes are known as GTPase of the immunity-associated protein family member 4 and 8, which are of crucial importance for immune function and development of immune-related diseases (Heinonen, 2015), and *GIMAP* genes are downregulated in human Type 1 diabetes (Jailwala et al., 2009). It is therefore tempting to assume that immunocompetence is related to high FE particularly in Jersey cows.

That was supported by the functional enrichment analysis by the GSEA procedure, where the identified KEGG pathways for regulated genes were related to primarily to immunocompetence functions (Leukocyte transendothelial migration, Primary immunodeficiency, Retinol metabolism, and Metbolism of xenobiotics by cytochrome P450) in Jersey cows. This is in agreement with studies in beef cattle, where Alexandre et al. (2015) reported a positive correlation between liver periportal lesions and functional enrichment for inflammatory response related functions in low-FE Nellore cattle.

Although the function of the DEGs for Holstein cows were primarily linked to aspects of lipid metabolism, the functional enrichment analysis applied to the whole gene set revealed that the top significant pathways for high-RFI (low FE) cows also related to immunocompetence functions in this breed (e.g. Primary immunodeficiency, Natural killer cell mediated cytotoxicity, Leukocyte transendothelial migration, Chemokine signalling pathway, FC gamma R mediated phagocytosis).

Therefore, the present project underlines the importance of immune related functions to obtain high FE and hence better economy efficiency. The liver plays a key role in this context, as the

responsible organ for acute phase reactions of the organism in addition to involvement in adaptive/specific immune responses.

Weighted gene co-expression network analysis (WGCNA)

The differential gene expression analysis only accounts for differences for individual genes, and it does not give a holistic overview about the coordinated function across genes to influence the biology underlying the trait of interest. The co-expression network analysis is applied across samples in a treatment group and can reveal, which genes have the same pattern of regulations (van Dam et al., 2017). Moreover, the co-expression network analysis avoids the multiple testing problems and subsequently increases the power of the analysis.

The rationale underlying this type of analysis is that genes, which are up and down regulated together (co-expressed) in relation to a specific biological function or mechanism also participate in the regulation of that biological mechanism/function. Therefore, when expression of groups of genes are correlated with RFI, it is possible that the biological mechanisms they are involved in also is of primary importance in the regulation of RFI and hence FE.

Thus, the information about the functional enrichment from the analysis of co-expressed genes provides a better understanding of the mechanisms, which are important determinants of RFI in Holstein and Jersey cows. The WGCNA considers the co-variation between gene expression and the trait if interest (here: RFI), and provide additional information to entangle the regulatory genes in the complex system (Ramayo-Caldas et al., 2018).

Therefore, the co-expression network analysis is important in order to investigate the groups of genes that could be key regulators in biological pathways underlying complex multi-factorial traits, such as RFI. Thus, the functional enrichment confirmed and underlined the breed specific traits relating to RFI, by highlighting the importance in this respect of lipid and cholesterol metabolism in Holstein and immune related function in Jersey.

Somewhat surprisingly, we could not find any biologically meaningful modules associated with differences in dietary exposure, although all absorbed nutrients (except for lipids) are transported directly to the liver and exposed to possible post-absorptive hepatic modifications, before being distributed to other tissues in the body.

Hepatic adaptation to differences in patterns of absorbed nutrients does therefore not appear to important in relation to RFI, and has therefore only been given little attention in this thesis in relation to identifying of regulatory genes. For breeding purposes, this may however be relevant to know for determination of expression patterns of candidate genes identified in this PhD project, it is not a major issue how the sampled cows are fed in practice.

As already pointed out, genetic markers for cholesterol and lipid metabolic pathways were identified as being positively associated with RFI in the Holstein cows. In pigs gene expression profile, cholesterol and lipid metabolism have also been associated with the FE trait, since rate of lipogenesis and steroidogenesis in the liver and fat tissues were closely related to FE (Lkhagvadorj et al., 2010). In another pigs study by Y. Zhao et al. (2016) key genes involved in steroid hormone metabolism (*CYP11A1, HSD17B2* and *UGT2B4*) were upregulated in liver of the high-FE groups. In addition, Y. Chen et al. (2011) reported DE genes between high and low RFI Angus cattle, which lipid and carbohydrate metabolism in addition to cellular growth and proliferation, cellular assembly and organization, cell signalling, drug metabolism and protein synthesis.

In Jersey, the list of significant pathways distinguishing high- and low-RFI cows was related to immune system functions. Many other studies suggested have highlighted a close association between FE and the immune system. Recently, a gene co-expression networks study in pigs by Ramayo-Caldas et al. (2018) revealed that the significant modules associated with RFI trait involved in inflammation, immune response, lipid metabolism as well as thermoregulation. In another study, immune response, response to oxidative stress and protein metabolism were the essential biological pathways that differentiated high- and low-RFI pigs (Gondret et al., 2017). However, no previous studies have specifically focused on these aspects in dairy cattle, but high FE appears to involve many similar biological functions across species and breeds. A hygienic environment in the barn has positive effects on the FE of animals (Johnson, 2012; Mpetile, 2014; Van Eerden et al., 2004; Vigors et al., 2016), but Jersey may have adopted special strategies compared to Holstein to cope with this, which appears to agree with the general conception of a more robust cow.

Since the identified modules were biologically meaningful and represent functions associated with feed efficiency in other species and breeds, it is proposed that the identified central (hub) genes in the module networks can be breed specific candidate genes for FE in dairy cattle.

The upstream regulator analysis from the IPA identifies potential upstream molecules (regulators) that can explain expression changes in the datasets (Krämer et al., 2013). This analysis is important in order to predict the direction (activation or inhibition) of changes in the dataset (Krämer et al., 2013). We identified ATP7B as a top upstream regulator for the salmon module. This protein regulates copper transport in and out of liver cells, and the elimination of excess copper in bile, using energy from the molecule adenosine triphosphate (ATP). It also appears to be involved in the excretion of copper into milk. As an important co-factor in many enzymes, copper has long been used as a dietary additive to improve growth performance and

feed efficiency of e.g. pigs (Blaabjerg & Poulsen, 2017). Although it is not straightforward, the differential expression of the gene may reflect differences in ability to metabolise and deposit fat in the body (Huster et al., 2007), and ATP7B appears to be activated in high RFI (low FE).

In the present study, cholesterol synthesis was activated in the IPA upstream regulator analysis. Furthermore, the activation of lipid metabolism in the disease function analysis supports the evidence from the GO term and pathway analyses. As lipid and cholesterol metabolism, and fat synthesis in particular, are activated in the high RFI group, we can assume that the high RFI group was inefficient in directing nutrients involved in fat metabolic pathways towards milk synthesis. Hence, animals with high RFI (low FE) have been found to have higher levels of fat in the body (E. C. Richardson et al., 2004), which is consistent with Arthur et al. (2001), who reported a positive relationship between RFI and average back fat in beef carcasses.

The top upstream regulator in Jersey cows was Interferon Gamma (IFNG), which has an interesting relationship to interactions among nutrition, metabolism, and the immune system (Schroder et al., 2004). This gene encodes a soluble cytokine that is a member of the type II interferon class. IFNG was predicted to be inhibited in high RFI Jersey cows. This protein is secreted from cells of both the innate and adaptive immune systems and plays an important role in regulating immune function in animals. IFNG is important, because it directly inhibits viral replication. The down-regulation of this cytokine in the high RFI group in Jersey cows might affect feed efficiency indirectly due to reduced immunocompetence. Another interesting upstream regulator in Jersey cows was IL10RA (Interleukin 10 Receptor Subunit Alpha), which was predicted to be activated in cows with high-RFI. IL10RA is a receptor with anti-inflammatory properties (Gondret et al., 2016). The activation of this gene might result in inhibition of the synthesis of pro-inflammatory cytokines. Reynolds et al. (2017) reported that IL10RA was differentially expressed in rumen papillae of divergent average daily gain steers, and these authors showed a negative association between the inflammatory response and feed FE. Thus, the activation of IL10RA in the high RFI group would reflect depressed inflammatory response in Jersey cows.

In Holstein, module that affected by changes in the forage:concentrate ratio in the diet is the top significant module, which could be related to biologically meaningful pathways in the functional pathways analysis, namely the Magenta module, and it was enriched for Triglyceride homeostasis process. In ruminants, triglycerides in adipose tisuses is the major energy reserve in the body, and fatty acids can be mobilized from these reserves to sustain milk fat synthesis during early lactation (Drackley, 2007). Feeding diets to dairy cows with either 30 or 50% concentrate in dietary DM during the first 10 weeks of lactation had marked influence on total

dietary energy intake, milk yield, body weight changes and metabolite patterns reflecting fat mobilization (Reist et al., 2003), and such changes in the pattern of nutrient absorption can onviously be related to changes in gene expression associated with lipid metabolism processes.

To sum up, in the co-expression analysis it was observed that Holstein co-regulatory genes were centred around processes affecting lipid and cholesterol metabolism compared to Jersey, where the main biological functions affected related to immunocompetence and inflammatory responses.

Integrative genomics analysis / eQTL mapping

The integration of transcriptomics data with genomic data was the final approach in this PhD project to gain additional information about regulatory regions responsible for expression of candidate genes determining RFI and FE. The hypothesis underlying the integrative genomics analysis is that SNPs associated with the expression of candidate genes are involved in, or in linkage with, genomic regions regulating their expression.

Different strategies were applied to identify the eQTLs that associated with expression of the RFI candidate genes. For this the two data sets for cows fed the C diet (1) and HC diet (2) were used. Since two different set of gene expression data were used in this approach from the same cows (fed either the C or HC diets) against the same genotype data, this analyses could be a interpreted as a replicate analysis. It was confirmed that certain regions (eQTLs) appear as significantly associated with the targeting genes (RFI's candidate genes) in both dataset analyses (C and HC). This can explain that the regions that were found significant were highly associated with the expression of the candidate genes (DEGs and HGs).

When we compared the results of the SNPs locations identified in the present study with previously reported QTLs and variants from GWAS studies reported in the Animal genome cattle QTLdatabase, we identified several overlaps of our eQTLs with QTLs from previous studies. The QTLs overlapping with our eQTLs were associated with a different type of traits, such as RFI, rump width, metabolic body weight, body weight gain, body weight (yearling), body weight, body depth, average daily gain as well as average daily feed intake.

Most of the gene functions and molecular pathways have been extensively elaborated in paper 1 and paper 2. Therefore, in the eQTL analysis, we focused more on the overlapping of the SNPs with previous GWAS study (genomic region).

The eQTLs that associated the most with several traits were associated with the expression of *ELOVL6* and *FDXR* genes. Only the *GIMAP4* gene was previously associated to production traits, which were RFI, rump witdh, metabolic body weight, body weight gain, body weight

(yearling), body weight, body depth, average daily gain as well as average daily feed intake (Lu et al., 2013; Snelling et al., 2010). However, the same chromosomal region contains QTLs for other traits, such as reproduction, milk, meat and carcass characteristics, health and exterior association traits (Cole et al., 2011; Lindholm-Perry et al., 2012).

All the associations with different types of traits found within this 1Mb region shows that the significant region might have control points for several targeting genes. The eQTLs identified are close to the QTL for production traits and for FE traits. At the same time this confirms the biological meaning related to FE trait of our findings.

The integrative genomics, also known as eQTL mapping, would bring us to another level of validation of the candidate genes that were identified in DEG and CEG analyses.

Moreover, the expression level of the targeted genes across animals with different genotype was correlated with the FE group. Therefore, these eQTLs are potential genetic markers to be tested for selection of high FE cows by using the favourable genotype combinations.

To identify precisely the location of the regulatory regions, the same analysis should be performed with a larger sample size to validate the findings and identified candidate genes and eQTLs. The findings of the present study suggest that the genomic region around the SNP markers can be used as potential biomarkers for feed efficiency and used to predict feed efficiency level of the animal. There is a risk that selection of dairy animals for feed efficiency based on certain traits or markers could also jeopardize fertility and other production trait due to unfavourable associations. Hence, all possible tests and validations need to be considered before actual application.

Breed differences and implication for improving feed efficiency through breeding

Blake et al. (1986) found there was no differences between the Holstein and Jersey breed in terms of efficiencies of utilizing the dry matter. However, it has also been reported that Holstein and Jersey cows differ in milk composition and DMI (Shetty et al., 2017). None of the DEG identified in the present study was in common across the two breeds, but when the DEGs were assigned to the functional enrichment analysis, it was that within each breed, the identified DEGs were quite consistently involved in similar biological pathways or functions.

We discovered more genes that were DE in Holstein compared to Jersey. This could in part be because of the variations between the two RFI groups (high and low) was weaker in the Jerseys.

The analysis by integrative genomics approach also confirmed that the two breeds (Holstein and Jersey) were different from the transcriptomic and a genomics point of view with respect to physiological and biological mechanisms associated with RFI. In the Holstein breed, DEGs,

CEGs, HGs and upstream regulators were involved in lipid, fat and cholesterol metabolisms. While, the Jersey breed DEGs, CEGs, HGs as well as the upstream regulators were mostly involved in immunocompetence related.

Conclusions

Feed efficiency is a complex trait that needs to be understood for further improvements. The biology and genetic foundation for this trait has been investigated and studied for a long time, however, due to difficulties, among other things, to accurately estimate FE in group fed cows and complex changes in regression coefficients between phenotypic traits used to estimate FE, it is difficult to include this economically important trait in breeding programmes. The transcriptomics and new systems biology approaches applied in the present PhD study, can contribute to significant new knowledge and improve our understanding of the biological mechanisms underlying FE in dairy cattle. Potentially overlooked new biomarkers for this trait can be identified, and especially high throughput data will enable us to better understand the complexity of the underlying systems biology. In the present study, it was possible to identify candidate genes (DEG and hub genes) as well as eQTLs RFI, and the study points to breed specific biological strategies or phenotypic traits associated with low RFI. Due to the rather small sample size in this PhD project, it is obvious that the findings need to be validated and confirmed in larger populations of cows order before biomarkers can be identified as reliable and taken into use in breeding programmes.

The candidate genes that could be prioritized for the development of biomarkers for low RFI and FE in Holstein cattle are *ACACA*, *CYP11A1, CYP7A1, ELOVL6, DGAT2* and *BDH2*, while for Jersey they are *FDXR, GIMAP4* and *GIMAP8*. The functional analysis revealed that these genetic markers for high-RFI (poor FE) in Holstein cows were associated with functional traits relating to cholesterol, ketone body and lipid metabolism. Future studies are needed to reveal, whether this reflects functional differences in lipid metabolic pathways exclusively in the liver (from where biopsies were sampled), or also in the mammary gland (milk fat synthesis and excretory capacity) and adipose tissues. In Jersey cows, however, the genes enriched in the high-RFI group were all associated with immunocompetence related functions.

These candidate genes could, subject to validation in larger populations, be prioritized for the development of biomarkers, and the genetic variants could be used in new genomic selection methods that include biological or functional information on SNPs.

To the best of the author's knowledge, the present PhD study is the first attempt to search for and evaluate the utility of candidate genes that could be included in genomic selection for low RFI (high FE) based on mRNA and eQTL information on the RFI trait in dairy cattle. Likewise, the breed specific functional changes and completely different biological strategies to attain high FE have never previously been described either. The present study therefore provides additional new

and useful information to be used for the improvement of the FE trait in breeding programmes in the future. Therefore, the present study will bring us a step further ahead and add more information that can be used in the future to genetically improve feed efficiency trait, genetically.

Future perspectives

The present study has revealed that omics techniques and the new systems biology techniques can be used to unravel the complex biology underlying the FE trait in dairy cattle, identify new candidate genes and SNPs that potentially can be integrated in future breeding programs. Two immediate future types of studies should be initiated to follow up on these findings.

Firstly, additional test(s) using different sets or larger populations of dairy cows should be conducted to validate the findings from the present PhD project. A simple validation test to confirm the findings (candidate genes), would be to test expression of specific genes using the RT-qPCR (Real Time-Quantitative Polymerase Chain Reaction) technique in a different and larger population, and comparing expression patterns in other tissues, particularly adipose and the mammary gland. Since mammary biopsying is associated with risk of mammary infections, it would be relevant to test, if mammary cells shedded in milk could be used to reveal the relevant changes in gene expression.

Secondly, genetic data should be integrated with metabolomics data in order to obtain important additional information needed for more accurate identification of the best candidate genes and pathways associated with the manifestation of the complex FE trait in dairy cattle. The combination of different layers of omics data together with knowledge about the phenotypes would contribute to a holistic approach towards genetic improvement of this complex trait.

The identification of the candidate genes through this type of analysis should also be conducted in other more easily accessible cells or tissues from blood, milk, muscle or adipose tissue. Particularly with respect to lipid metabolism associated candidate genes in Holstein cows, it is important to establish, whether the finding in this PhD project based on liver tissue also apply to the most important lipogenic tissues, namely the mammary gland and adipose tissues, or whether other appropriate biomarkers should be considered for these types of cells or tissues. Also, histological analysis of liver tissues and biochemical analysis of blood metabolites of inflammatory markers could be conducted to validate the implication of the biological pathways identified in the PhD project. The focus should particularly be related to inflammatory response and immunocompetence traits in Jersey cows and lipid and cholesterol metabolism in Holsteins.

For application in future breeding programmes, analyses for biomarker expression should be based on easily accessible tissue(s) or cell(s).

Finally, along with the FE trait, other traits such as fertility, health and reproduction must be considered very carefully in order to ensure that selection based on identified FE candidate genes

will not be associated with negative effects on other important traits for economical sustainability of livestock production.

References

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., . . . Moreno, R. F. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science, 252*(5013), 1651-1656.

Alexandre, P. A., Kogelman, L. J., Santana, M. H., Passarelli, D., Pulz, L. H., Fantinato-Neto, P., . . . Coutinho, L. L. (2015). Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. *BMC genomics, 16*(1), 1.

Allen, M. S., & Bradford, B. (2009). *Nutritional control of feed intake in dairy cattle.* Paper presented at the Florida Nutrition Conference.

Alter, O., Brown, P. O., & Botstein, D. (2003). Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proceedings of the National Academy of Sciences, 100*(6), 3351-3356.

Amar, D., Safer, H., & Shamir, R. (2013). Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS computational biology, 9*(3), e1002955.

Anders, S., Pyl, P. T., & Huber, W. (2014). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, btu638.

Arthur, P., Archer, J., Johnston, D., Herd, R., Richardson, E., & Parnell, P. (2001). Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *Journal of animal science, 79*(11), 2805-2811.

Astessiano, A. L., Meikle, A., Fajardo, M., Gil, J., Mattiauda, D. A., Chilibroste, P., & Carriquiry, M. (2015). Metabolic and endocrine profiles and hepatic gene expression of Holstein cows fed total mixed ration or pasture with different grazing strategies during early lactation. *Acta Veterinaria Scandinavica, 57*(1), 70.

Baird, G. D. (1982). Primary ketosis in the high-producing dairy cow: clinical and subclinical disorders, treatment, prevention, and outlook. *Journal of Dairy Science, 65*(1), 1-10.

Ballouz, S., Verleyen, W., & Gillis, J. (2015). Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics, 31*(13), 2123-2130.

Bauer, M. (2007). RNA in forensic science. *Forensic Science International: Genetics, 1*(1), 69-74.

Baumgard, L., Collier, R., & Bauman, D. (2017). A 100-Year Review: Regulation of nutrient partitioning to support lactation. *Journal of dairy science, 100*(12), 10353-10366.

Bergman, E. (1971). Hyperketonemia-Ketogenesis and Ketone Body Metabolism2. *Journal of Dairy Science, 54*(6), 936-948.

Berry, D., & Crowley, J. (2013). Cell biology symposium: genetics of feed efficiency in dairy and beef cattle. *Journal of animal science, 91*(4), 1594-1613.

Bilotta, G., Brazier, R., & Haygarth, P. (2007). The impacts of grazing animals on the quality of soils, vegetation, and surface waters in intensively managed grasslands. *Advances in agronomy, 94*, 237-280.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., . . . Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics, 25*(8), 1091-1093.

Bioinformatics, B. (2011). FastQC: a quality control tool for high throughput sequence data. *Cambridge, UK: Babraham Institute*.

Bionaz, M. (2014). Nutrigenomics Approaches to Fine-Tune Metabolism and Milk Production: Is This the Future of Ruminant Nutrition? *Advances in Dairy Research, 2014*.

Bionaz, M., & Loor, J. J. (2008). Gene networks driving bovine milk fat synthesis during the lactation cycle. *BMC genomics, 9*(1), 366.

Bionaz, M., & Loor, J. J. (2012). Ruminant metabolic systems biology: reconstruction and integration of transcriptome dynamics underlying functional responses of tissues to nutrition and physiological state. *Gene regulation and systems biology, 6*, GRSB. S9852.

Blake, R. W., Custodio, A. A., & Howard, W. H. (1986). Comparative Feed Efficiency of Holstein and Jersey Cows1. *Journal of Dairy Science, 69*(5), 1302-1308.

Blaabjerg, K., & Poulsen, H. D. (2017). The use of zinc and copper in pig production.

Breitling, R. (2010). What is systems biology? *Frontiers in physiology, 1*, 159.

Bu, D., Bionaz, M., Wang, M., Nan, X., Ma, L., & Wang, J. (2017). Transcriptome difference and potential crosstalk between liver and mammary tissue in mid-lactation primiparous dairy cows. *PloS one, 12*(3), e0173082.

Buermans, H., & Den Dunnen, J. (2014). Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1842*(10), 1932-1941.

Casper, D. P. (2008). *Factors affecting feed efficiency of dairy cows*. Paper presented at the Proceedings of the Tri-State Dairy Nutrition Conference'.(Ed. ML Eastridge) pp.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience, 4*(1), 7.

Chen, H. C., & Farese, R. V. (2000). DGAT and triglyceride synthesis: a new target for obesity treatment? *Trends in cardiovascular medicine, 10*(5), 188-192.

Chen, S., He, H., & Liu, X. (2017). Tissue expression profiles and transcriptional regulation of elongase of very long chain fatty acid 6 in bovine mammary epithelial cells. *PloS one, 12*(4), e0175777.

Chen, Y., Gondro, C., Quinn, K., Herd, R., Parnell, P., & Vanselow, B. (2011). Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. *Animal genetics, 42*(5), 475-490.

Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, T. J., Crooker, B. A., . . . Matukumalli, L. K. (2011). Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC genomics, 12*(1), 408.

Coleman, J., Berry, D., Pierce, K., Brennan, A., & Horan, B. (2010). Dry matter intake and feed efficiency profiles of 3 genotypes of Holstein-Friesian within pasture-based systems of milk production. *Journal of Dairy Science, 93*(9), 4318-4331.

Connor, E. (2015). Invited review: Improving feed efficiency in dairy production: Challenges and possibilities. *animal, 9*(3), 395-408.

Connor, E., Hutchison, J., Norman, H., Olson, K., Van Tassell, C., Leith, J., & Baldwin, R. (2013). Use of residual feed intake in Holsteins during early lactation shows potential to improve feed efficiency through genetic selection. *Journal of animal science, 91*(8), 3978-3988.

Connor, E. E., Hutchison, J. L., & Norman, H. D. (2012). Estimating feed efficiency of lactating dairy cattle using residual feed intake. *Feed Efficiency in the Beef Industry*, 159-173.

Council, N. R. (2001). *Nutrient requirements of dairy cattle: 2001*: National Academies Press.

Cummins, S. B., Waters, S. M., Evans, A., Lonergan, P., & Butler, S. T. (2012). Genetic merit for fertility traits in Holstein cows: III. Hepatic expression of somatotropic axis genes during pregnancy and lactation. *Journal of Dairy Science, 95*(7), 3711-3721.

Dai, W., Chen, Q., Wang, Q., White, R. R., Liu, J., & Liu, H. (2017). Complementary transcriptomic and proteomic analyses reveal regulatory mechanisms of milk protein production in dairy cows consuming different forages. *Scientific reports, 7*, 44234.

Dann, H., Litherland, N., Underwood, J., Bionaz, M., D'angelo, A., McFadden, J., & Drackley, J. (2006). Diets During Far-Off and Close-Up Dry Periods Affect Periparturient Metabolism and Lactation in Multiparous Cows1. *Journal of Dairy Science, 89*(9), 3563-3577.

Debnath, M., Prasad, G. B., & Bisen, P. S. (2010). *Molecular diagnostics: promises and possibilities*: Springer Science & Business Media.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21.

Drackley, J. K. (2007). Overview of fat digestion and metabolism in dairy cows.

Drackley, J. K., Overton, T. R., & Douglas, G. N. (2001). Adaptations of glucose and long-chain fatty acid metabolism in liver of dairy cows during the periparturient period. *Journal of Dairy Science, 84*, E100-E112.

Dündar, F., Skrabanek, L., & Zumbo, P. (2015). Introduction to differential gene expression analysis using RNA-seq. *Appl. Bioinformatics*, 1-67.

Fang, X., Zhang, J., Xu, H., Zhang, C., Du, Y., Shi, X., . . . Lan, X. (2012). Polymorphisms of diacylglycerol acyltransferase 2 gene and their relationship with growth traits in goats. *Molecular biology reports, 39*(2), 1801-1807.

Felippe, E. W., Ivan, G., & Neto, A. T. (2017). *Comparison of Holstein X Jersey crossbred cows with straightbreds for productive and reproductive efficiency* (Vol. 22).

Finotello, F., & Di Camillo, B. (2015). Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Briefings in functional genomics, 14*(2), 130-142.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., . . . Von Mering, C. (2012). STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research, 41*(D1), D808-D815.

Gilmore, L. O. (1952). *Dairy cattle breeding*: JB Lippincott Co; Chicago.

Gleeson, S., Mulroy, E., & Clarke, D. E. (2016). Lactation ketoacidosis: an unusual entity and a review of the literature. *The Permanente Journal, 20*(2), 71.

Gondret, F., Vincent, A., Houée-Bigot, M., Siegel, A., Lagarrigue, S., Causeur, D., . . . Louveau, I. (2017). A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs. *BMC genomics, 18*(1), 244.

Gondret, F., Vincent, A., Houée-Bigot, M., Siegel, A., Lagarrigue, S., Louveau, I., & Causeur, D. (2016). Molecular alterations induced by a high-fat high-fiber diet in porcine adipose tissues: variations according to the anatomical fat location. *BMC genomics, 17*(1), 120.

Grant, G. R., Farkas, M. H., Pizarro, A. D., Lahens, N. F., Schug, J., Brunk, B. P., . . . Pierce, E. A. (2011). Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics, 27*(18), 2518-2528.

Griebel, T., Zacher, B., Ribeca, P., Raineri, E., Lacroix, V., Guigó, R., & Sammeth, M. (2012). Modelling and simulating generic RNA-Seq experiments with the flux simulator. *Nucleic acids research, 40*(20), 10073-10083.

Ha, M. J., Baladandayuthapani, V., & Do, K.-A. (2015). DINGO: differential network analysis in genomics. *Bioinformatics, 31*(21), 3413-3420.

Heinonen, M. (2015). Players of Th-cell differentiation and immune dysregulation–focus on GIMAP family genes.

Herd, R., & Arthur, P. (2009). Physiological basis for residual feed intake 1. *Journal of animal science, 87*(14_suppl), E64-E71.

Herdt, T. H. (2000). Ruminant adaptation to negative energy balance: Influences on the etiology of ketosis and fatty liver. *Veterinary Clinics: Food Animal Practice, 16*(2), 215-230.

Hoffman, K., DeClue, R., & Emmick, D. (2000). Prescribed grazing and feeding management of lactating dairy cows. *New York State Grazing Lands Conservation Initiative in Cooperation with the USDA-NRCS, Syracuse, New York*.

Horvath, S., & Langfelder, P. (2011). Tutorials for the WGCNA package for R: WGCNA Background and glossary.

Hu, Z.-L., Park, C. A., & Reecy, J. M. (2015). Developmental progress and current status of the Animal QTLdb. *Nucleic acids research, 44*(D1), D827-D833.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols, 4*(1), 44.

Hubner, N., Wallace, C. A., Zimdahl, H., Petretto, E., Schulz, H., Maciver, F., . . . Zidek, V. (2005). Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature genetics, 37*(3), 243.

Hung, J.-H. (2013). Gene set/pathway enrichment analysis *Data Mining for Systems Biology* (pp. 201-213): Springer.

Huster, D., Purnat, T. D., Burkhead, J. L., Ralle, M., Fiehn, O., Stuckert, F., . . . Lutsenko, S. (2007). High copper selectively alters lipid metabolism and cell cycle machinery in the mouse model of Wilson disease. *Journal of Biological Chemistry, 282*(11), 8343-8355.

Jailwala, P., Waukau, J., Glisic, S., Jana, S., Ehlenbach, S., Hessner, M., . . . Wang, X. (2009). Apoptosis of CD4+ CD25high T cells in type 1 diabetes may be partially mediated by IL-2 deprivation. *PloS one, 4*(8), e6527.

Jensen, D., Bessesen, D., Etienne, J., Eckel, R., & Neville, M. (1991). Distribution and source of lipoprotein lipase in mouse mammary gland. *Journal of Lipid Research, 32*(5), 733-742.

Jing, L., Hou, Y., Wu, H., Miao, Y., Li, X., Cao, J., . . . Zhao, S. (2015). Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an important network for differential Residual Feed Intake in pigs. *Scientific reports, 5*.

Johnson, R. (2012). Fueling the immune response: What's the cost? *Feed efficiency in swine* (pp. 211-223): Springer.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research, 28*(1), 27-30.

Kitano, H. (2002). Systems biology: a brief overview. *Science, 295*(5560), 1662-1664.

Knapp, J., Laur, G., Vadas, P., Weiss, W., & Tricarico, J. (2014). Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. *Journal of Dairy Science, 97*(6), 3231-3261.

Koch, R. M., Swiger, L. A., Chambers, D., & Gregory, K. E. (1963). Efficiency of Feed Use in Beef Cattle. *J Anim Sci, 22*.

Krämer, A., Green, J., Pollard Jr, J., & Tugendreich, S. (2013). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics, 30*(4), 523-530.

Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor protocols, 2015*(11), pdb. top084970.

Laflamme, L. B. (1973). Factors affecting feed intake and feed efficiency of lactating dairy cows.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics, 9*. doi:10.1186/1471-2105-9-559

Lawrence, J., Mintert, J., Anderson, J. D., & Anderson, D. P. (2008). Feed grains and livestock: Impacts on meat supplies and prices. *Choices, 23*(2), 11.

Le Mignon, G., Désert, C., Pitel, F., Leroux, S., Demeure, O., Guernec, G., . . . Lagarrigue, S. (2009). Using transcriptome profiling to characterize QTL regions on chicken chromosome 5. *BMC genomics, 10*(1), 575.

Lee, J., Karnuah, A. B., Rekaya, R., Anthony, N. B., & Aggrey, S. E. (2015). Transcriptomic analysis to elucidate the molecular mechanisms that underlie feed efficiency in meat-type chickens. *Molecular genetics and genomics, 290*(5), 1673-1682.

Li, B., Berglund, B., Fikse, W., Lassen, J., Lidauer, M. H., Mäntysaari, P., & Løvendahl, P. (2017). Neglect of lactation stage leads to naive assessment of residual feed intake in dairy cattle. *Journal of Dairy Science, 100*(11), 9076-9084.

Li, B., Fikse, W., Lassen, J., Lidauer, M., Løvendahl, P., Mäntysaari, P., & Berglund, B. (2016). Genetic parameters for dry matter intake in primiparous Holstein, Nordic Red, and Jersey cows in the first half of lactation. *Journal of Dairy Science*.

Li, J., Xu, X., Zhang, Q., Wang, X., Deng, G., Fang, X., . . . Xu, S. (2009). Association between single nucleotide polymorphisms in the Dgat2 gene and beef carcass and quality traits in commercial feedlot steers. *Asian-Australasian Journal of Animal Sciences, 22*(7), 943-954.

Li, W., Bu, D., Wang, J., Nan, X., Sun, P., & Zhou, L. (2013). Effect of two different diets on liver gene expression associated with glucose metabolism in dairy cows. *Livestock Science, 158*(1), 223-229.

Lindholm-Perry, A. K., Kuehn, L. A., Rempel, L. A., Smith, T. P., Cushman, R. A., McDaneld, T. G., . . . Freetly, H. C. (2012). Evaluation of Bovine chemerin (RARRES2) Gene Variation on Beef Cattle Production Traits1. *Frontiers in genetics, 3*, 39.

Lkhagvadorj, S., Qu, L., Cai, W., Couture, O. P., Barb, C. R., Hausman, G. J., . . . Tuggle, C. K. (2009). Gene expression profiling of the short-term adaptive response to acute caloric restriction in liver and adipose tissues of pigs differing in feed efficiency. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 298*(2), R494-R507.

Lkhagvadorj, S., Qu, L., Cai, W., Couture, O. P., Barb, C. R., Hausman, G. J., . . . Tuggle, C. K. (2010). Gene expression profiling of the short-term adaptive response to acute caloric restriction in liver and adipose tissues of pigs differing in feed efficiency. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 298*(2), R494-R507.

Loor, J. J., Everts, R. E., Bionaz, M., Dann, H. M., Morin, D. E., Oliveira, R., . . . Lewin, H. A. (2007). Nutrition-induced ketosis alters metabolic and signaling gene networks in liver of periparturient dairy cows. *Physiological genomics, 32*(1), 105-116.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology, 15*(12), 1.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS computational biology, 13*(5), e1005457.

Lu, D., Miller, S., Sargolzaei, M., Kelly, M., Vander Voort, G., Caldwell, T., . . . Moore, S. (2013). Genome-wide association analyses for growth and feed efficiency traits in beef cattle. *Journal of animal science, 91*(8), 3612-3633.

Madsen, J., Bjerg, B. S., Hvelplund, T., Weisbjerg, M. R., & Lund, P. (2010). Methane and carbon dioxide ratio in excreted air for quantification of the methane production from ruminants. *Livestock Science, 129*(1), 223-227.

Matsuzaka, T., Atsumi, A., Matsumori, R., Nie, T., Shinozaki, H., Suzuki-Kemuriyama, N., . . . Shimada, M. (2012). Elovl6 promotes nonalcoholic steatohepatitis. *Hepatology, 56*(6), 2199-2208.

Maza, E. (2016). In Papyro comparison of TMM (edgeR), RLE (DESeq2), and MRN normalization methods for a simple two-conditions-without-replicates RNA-Seq experimental design. *Frontiers in genetics, 7*, 164.

Mazzoni, G., & Kadarmideen, H. N. (2016). Computational Methods for Quality Check, Preprocessing and Normalization of RNA-Seq Data for Systems Biology and Analysis *Systems Biology in Animal Production and Health, Vol. 2* (pp. 61-77): Springer.

Mazzoni, G., Kogelman, L. J., Suravajhala, P., & Kadarmideen, H. N. (2015). Systems genetics of complex diseases using RNA-sequencing methods. *International Journal of Bioscience, Biochemistry and Bioinformatics, 5*(4), 264.

McCabe, M., Waters, S., Morris, D., Kenny, D., Lynn, D., & Creevey, C. (2012). RNA-seq analysis of differential gene expression in liver from lactating dairy cows divergent in negative energy balance. *BMC genomics, 13*(1), 193.

Minuti, A., Zhou, Z., Graugnard, D. E., Rodriguez-Zas, S. L., Palladino, A. R., Cardoso, F. C., . . . Loor, J. J. (2015). Acute mammary and liver transcriptome responses after an intramammary Escherichia coli lipopolysaccharide challenge in postpartal dairy cows. *Physiological reports, 3*(4).

Moore, S., Mujibi, F., & Sherman, E. (2009). Molecular basis for residual feed intake in beef cattle 1. *Journal of animal science, 87*(14_suppl), E41-E47.

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., . . . Laurila, E. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics, 34*(3), 267-273.

Moran, B., Cummins, S. B., Creevey, C. J., & Butler, S. T. (2016). Transcriptomics of liver and muscle in Holstein cows genetically divergent for fertility highlight differences in nutrient partitioning and inflammation processes. *BMC genomics, 17*(1), 603.

Mpetile, Z. (2014). Effect of divergent selection for residual feed intake on immune system of Yorkshire pigs.

Nica, A. C., & Dermitzakis, E. T. (2013). Expression quantitative trait loci: present and future. *Phil. Trans. R. Soc. B, 368*(1620), 20120362.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2015). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics, 32*(2), 292-294.

Paradis, F., Yue, S., Grant, J. R., Stothard, P., Basarab, J. A., & Fitzsimmons, C. (2015). Transcriptomic analysis by RNA sequencing reveals that hepatic interferon-induced genes may be associated with feed efficiency in beef heifers. *J Anim Sci, 93*. doi:10.2527/jas.2015-8975

Pech, C. M., Veerkamp, R., Calus, M., Zom, R., van Knegsel, A., Pryce, J., & De Haas, Y. (2014). Genetic parameters across lactation for feed intake, fat-and protein-corrected milk, and liveweight in first-parity Holstein cattle. *Journal of Dairy Science, 97*(9), 5851-5862.

Pontes, B., Giráldez, R., & Aguilar-Ruiz, J. S. (2015). Biclustering on expression data: A review. *Journal of biomedical informatics, 57*, 163-180.

Potts, S., Boerman, J., Lock, A., Allen, M., & VandeHaar, M. (2015). Residual feed intake is repeatable for lactating Holstein dairy cows fed high and low starch diets. *Journal of Dairy Science, 98*(7), 4735-4747.

Pryce, J., Arias, J., Bowman, P., Davis, S., Macdonald, K., Waghorn, G., . . . Hayes, B. (2012). Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. *Journal of Dairy Science, 95*(4), 2108-2119.

Pryce, J., Gonzalez-Recio, O., Nieuwhof, G., Wales, W., Coffey, M., Hayes, B., & Goddard, M. (2015). Hot topic: Definition and implementation of a breeding value for feed efficiency in dairy cows. *Journal of Dairy Science, 98*(10), 7340-7350.

Pryce, J., Wales, W. d., De Haas, Y., Veerkamp, R., & Hayes, B. (2014). Genomic selection for feed efficiency in dairy cattle. *animal, 8*(1), 1-10.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

Rahmatallah, Y., Emmert-Streib, F., & Glazko, G. (2013). Gene Sets Net Correlations Analysis (GSNCA): a multivariate differential coexpression test for gene sets. *Bioinformatics, 30*(3), 360-368.

Ram, P. T., Mendelsohn, J., & Mills, G. B. (2012). Bioinformatics and systems biology. *Molecular oncology, 6*(2), 147-154.

Ramayo-Caldas, Y., Ballester, M., Sánchez, J. P., González-Rodríguez, O., Revilla, M., Reyer, H., . . . Quintanilla, R. (2018). Integrative approach using liver and duodenum RNA-Seq data identifies candidate genes and pathways associated with feed efficiency in pigs. *Scientific reports, 8*(1), 558.

Reist, M., Erdin, D., Von Euw, D., Tschuemperlin, K., Delavaud, C., Chilliard, Y., . . . Zbinden, Y. (2003). Concentrate feeding strategy in lactating dairy cows: metabolic and endocrine changes with emphasis on leptin. *Acta Veterinaria Scandinavica, 44*(1), P30.

Reynolds, J., Foote, A., Freetly, H., Oliver, W., & Lindholm-Perry, A. (2017). Relationships between inflammation-and immunity-related transcript abundance in the rumen and jejunum of beef steers with divergent average daily gain. *Animal genetics, 48*(4), 447-449.

Richardson, E., & Herd, R. (2004). Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. *Australian Journal of Experimental Agriculture, 44*(5), 431-440.

Richardson, E. C., Herd, R. M., Archer, J. A., & Arthur, P. F. (2004). Metabolic differences in Angus steers divergently selected for residual feed intake. *Aust J Exp Agric, 44*. doi:10.1071/ea02219

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, gkv007.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Roche, J. R., Friggens, N. C., Kay, J. K., Fisher, M. W., Stafford, K. J., & Berry, D. P. (2009). Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. *Journal of dairy science, 92*(12), 5769-5801.

Ropka-Molik, K., Knapik, J., Pieszka, M., Szmatoła, T., & Piórkowska, K. (2017). Nutritional modification of SCD, ACACA and LPL gene expressions in different ovine tissues. *Archives Animal Breeding, 60*(3), 243-250.

RYK- Ydelseskontrol, https://www.landbrugsinfo.dk/Kvaeg/RYK/Sider/Avereagemilkyieldisnowcloseto10000kgprcowandyears.aspx, Accessed date: 12-06-2017.

Salleh, M., Mazzoni, G., Höglund, J., Olijhoek, D., Lund, P., Løvendahl, P., & Kadarmideen, H. (2017). RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high-and low-residual feed intake in Nordic dairy cattle. *BMC genomics, 18*(1), 258.

Santana, M. H., Utsunomiya, Y. T., Neves, H. H., Gomes, R. C., Garcia, J. F., Fukumasu, H., . . . Leme, P. R. (2014). Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. *BMC genetics, 15*(1), 21.

Schadt, E. E., Lamb, J., Yang, X., Zhu, J., Edwards, S., GuhaThakurta, D., . . . Zhang, C. (2005). An integrative genomics approach to infer causal associations between gene expression and disease. *Nature genetics, 37*(7), 710.

Schroder, K., Hertzog, P. J., Ravasi, T., & Hume, D. A. (2004). Interferon-γ: an overview of signals, mechanisms and functions. *Journal of leukocyte biology, 75*(2), 163-189.

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., . . . Owen-Hughes, T. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *Rna, 22*(6), 839-851.

Seal, C., & Reynolds, C. (1993). Nutritional implications of gastrointestinal and liver metabolism in ruminants. *Nutrition research reviews, 6*(1), 185-208.

Seyednasrollah, F., Laiho, A., & Elo, L. L. (2015). Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in bioinformatics, 16*(1), 59-70.

Shabalin, A. A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics, 28*(10), 1353-1358.

Shetty, N., Løvendahl, P., Lund, M., & Buitenhuis, A. (2017). Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk. *Journal of Dairy Science, 100*(1), 253-264.

Shi, H., Wu, M., Zhu, J., Zhang, C., Yao, D., Luo, J., & Loor, J. (2017). Fatty acid elongase 6 plays a role in the synthesis of long-chain fatty acids in goat mammary epithelial cells. *Journal of Dairy Science, 100*(6), 4987-4995.

Sidova, M., Tomankova, S., Abaffy, P., Kubista, M., & Sindelka, R. (2015). Effects of post-mortem and physical degradation on RNA integrity and quality. *Biomolecular detection and quantification, 5*, 3-9.

Snelling, W., Allan, M., Keele, J., Kuehn, L., Mcdaneld, T., Smith, T., . . . Bennett, G. (2010). Genome-wide association study of growth in crossbred beef cattle 1 2. *Journal of animal science, 88*(3), 837-848.

Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., Rosales, M., & de Haan, C. (2006). *Livestock's long shadow: environmental issues and options*: Food & Agriculture Org.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Lander, E. S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences, 102*(43), 15545-15550.

Suravajhala, P., Kogelman, L. J., & Kadarmideen, H. N. (2016). Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare. *Genetics Selection Evolution, 48*(1), 1.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . Tsafou, K. P. (2014). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, gku1003.

Saatchi, M., Garrick, D. J., & Taylor, J. (2014). Genome-Wide Association Study of Feed Efficiency in Beef Cattle. *Animal Industry Report, 660*(1), 20.

Tan, C. Y., Virtue, S., Murfitt, S., Roberts, L. D., Phua, Y. H., Dale, M., . . . Vidal-Puig, A. (2015). Adipose tissue fatty acid chain length and mono-unsaturation increases with obesity and insulin resistance. *Scientific Reports, 5*, 18366.

Tempelman, R., Spurlock, D., Coffey, M., Veerkamp, R., Armentano, L., Weigel, K., . . . Lu, Y. (2015). Heterogeneity in genetic and nongenetic variation and energy sink relationships for residual feed intake across research stations and countries. *Journal of Dairy Science, 98*(3), 2013-2026.

Tesson, B. M., Breitling, R., & Jansen, R. C. (2010). DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics, 11*(1), 497.

Tipney, H., & Hunter, L. (2010). An introduction to effective use of enrichment analysis software. *Human genomics, 4*(3), 202.

Tizioto, P. C., Coutinho, L. L., Oliveira, P. S., Cesar, A. S., Diniz, W. J., Lima, A. O., . . . Mourão, G. B. (2016). Gene expression differences in Longissimus muscle of Nelore steers genetically divergent for residual feed intake. *Scientific reports, 6*, 39493.

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics, 25*(9), 1105-1111.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., . . . Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology, 28*(5), 511.

Tuckey, R. C., Nguyen, M. N., Chen, J., Slominski, A. T., Baldisseri, D. M., Tieu, E. W., . . . Li, W. (2011). Human cytochrome P450scc (CYP11A1) catalyses epoxide formation with ergosterol. *Drug Metabolism and Disposition*, dmd. 111.042515.

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., & de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in bioinformatics*, bbw139.

Van Eerden, E., Van Den Brand, H., Parmentier, H., De Jong, M., & Kemp, B. (2004). Phenotypic selection for residual feed intake and its effect on humoral immune responses in growing layer hens. *Poultry science, 83*(9), 1602-1609.

VandeHaar, M., Armentano, L., Weigel, K., Spurlock, D., Tempelman, R., & Veerkamp, R. (2016). Harnessing the genetics of the modern dairy cow to continue improvements in feed efficiency1. *Journal of Dairy Science, 99*(6), 4941-4954.

Veerkamp, R., Emmans, G., Cromie, A., & Simm, G. (1995). Variance components for residual feed intake in dairy cows. *Livestock Production Science, 41*(2), 111-120.

Velez, J., & Donkin, S. (2005). Feed restriction induces pyruvate carboxylase but not phosphoenolpyruvate carboxykinase in dairy cows. *Journal of Dairy Science, 88*(8), 2938-2948.

Vigors, S., O'Doherty, J. V., Kelly, A. K., O'Shea, C. J., & Sweeney, T. (2016). The effect of divergence in feed efficiency on the intestinal microbiota and the intestinal immune response in both unchallenged and lipopolysaccharide challenged ileal and colonic explants. *PloS one, 11*(2), e0148145.

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., . . . Perou, C. M. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic acids research, 38*(18), e178-e178.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10*(1), 57-63.

Watson, M. (2006). CoXpress: differential co-expression in gene expression data. *BMC Bioinformatics, 7*(1), 509.

Weber, K. L., Welly, B. T., Van Eenennaam, A. L., Young, A. E., Porto-Neto, L. R., Reverter, A., & Rincon, G. (2016). Identification of gene networks for residual feed intake in angus cattle using genomic prediction and RNA-seq. *PloS one, 11*(3), e0152274.

White, R. R., & Hall, M. B. (2017). Nutritional and greenhouse gas impacts of removing animals from US agriculture. *Proceedings of the National Academy of Sciences*, 201707322.

Wickramasinghe, S., Cánovas, A., Rincón, G., & Medrano, J. F. (2014). RNA-sequencing: a tool to explore new frontiers in animal genetics. *Livestock Science, 166*, 206-216.

Wilkinson, J. (2011). Re-defining efficiency of feed use by livestock. *animal, 5*(7), 1014-1022.

Wong, N. P. (2012). *Fundamentals of dairy chemistry*: Springer Science & Business Media.

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics, 8*(3), 206.

Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics, 26*(7), 873-881.

Xi, Y., Wu, F., Zhao, D., Yang, Z., Li, L., Han, Z., & Wang, G. (2016). Biological mechanisms related to differences in residual feed intake in dairy cows. *animal, 10*(8), 1311-1318.

Xu, Z., Ji, C., Zhang, Y., Zhang, Z., Nie, Q., Xu, J., . . . Zhang, X. (2016). Combination analysis of genome-wide association and transcriptome sequencing of residual feed intake in quality chickens. *BMC genomics, 17*(1), 594.

Yi, G., Yuan, J., Bi, H., Yan, W., Yang, N., & Qu, L. (2015). In-Depth Duodenal Transcriptome Survey in Chickens with Divergent Feed Efficiency Using RNA-Seq. *PloS one, 10*(9), e0136765.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology, 11*(2), 1.

Zachary, J. F., & McGavin, M. D. (2013). *Pathologic Basis of Veterinary Disease-E-Book*: Elsevier Health Sciences.

Zang, L., Wang, Y., Sun, B., Zhang, X., Yang, C., Kang, L., . . . Jiang, Y. (2016). Identification of a 13bp indel polymorphism in the 3′-UTR of DGAT2 gene associated with backfat thickness and lean percentage in pigs. *Gene, 576*(2), 729-733.

Zarek, C., Lindholm-Perry, A., Kuehn, L., & Freetly, H. (2017). Differential expression of genes related to gain and intake in the liver of beef cattle. *BMC research notes, 10*(1), 1.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol, 4*.

Zhang, H., Ao, C., Song, L., & Zhang, X. (2015). Effects of different model diets on milk composition and expression of genes related to fatty acid synthesis in the mammary gland of lactating dairy goats. *Journal of Dairy Science, 98*(7), 4619-4628.

Zhang, R., Zhu, W., & Mao, S. (2016). High-concentrate feeding upregulates the expression of inflammation-related genes in the ruminal epithelium of dairy cattle. *Journal of animal science and biotechnology, 7*(1), 42.

Zhang, Z. H., Jhaveri, D. J., Marshall, V. M., Bauer, D. C., Edson, J., Narayanan, R. K., . . . Wray, N. R. (2014). A comparative study of techniques for differential expression analysis on RNA-Seq data. *PloS one, 9*(8), e103207.

Zhao, W., Langfelder, P., Fuller, T., Dong, J., Li, A., & Hovarth, S. (2010). Weighted gene coexpression network analysis: state of the art. *Journal of biopharmaceutical statistics, 20*(2), 281-300.

Zhao, Y., Hou, Y., Liu, F., Liu, A., Jing, L., Zhao, C., . . . Li, X. (2016). Transcriptome Analysis Reveals that Vitamin A Metabolism in the Liver Affects Feed Efficiency in Pigs. *G3: Genes| Genomes| Genetics, 6*(11), 3615-3624.

Accompanying Paper and Manuscripts

Paper 1: RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle

Salleh et al. BMC Genomics (2017) 18:258
DOI 10.1186/s12864-017-3622-9

BMC Genomics

RESEARCH ARTICLE Open Access

RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle

M. S. Salleh[1], G. Mazzoni[1], J. K. Höglund[2], D. W. Olijhoek[2,3], P. Lund[3], P. Løvendahl[2] and H. N. Kadarmideen[4*]

Abstract

Background: The selective breeding of cattle with high-feed efficiencies (FE) is an important goal of beef and dairy cattle producers. Global gene expression patterns in relevant tissues can be used to study the functions of genes that are potentially involved in regulating FE. In the present study, high-throughput RNA sequencing data of liver biopsies from 19 dairy cows were used to identify differentially expressed genes (DEGs) between high- and low-FE groups of cows (based on Residual Feed Intake or RFI). Subsequently, a profile of the pathways connecting the DEGs to FE was generated, and a list of candidate genes and biomarkers was derived for their potential inclusion in breeding programmes to improve FE.

Results: The bovine RNA-Seq gene expression data from the liver was analysed to identify DEGs and, subsequently, identify the molecular mechanisms, pathways and possible candidate biomarkers of feed efficiency. On average, 57 million reads (short reads or short mRNA sequences < ~200 bases) were sequenced, 52 million reads were mapped, and 24,616 known transcripts were quantified according to the bovine reference genome. A comparison of the high- and low-RFI groups revealed 70 and 19 significantly DEGs in Holstein and Jersey cows, respectively. The interaction analysis (high vs. low RFI x control vs. high concentrate diet) showed no interaction effects in the Holstein cows, while two genes showed interaction effects in the Jersey cows. The analyses showed that DEGs act through certain pathways to affect or regulate FE, including steroid hormone biosynthesis, retinol metabolism, starch and sucrose metabolism, ether lipid metabolism, arachidonic acid metabolism and drug metabolism cytochrome P450.

Conclusion: We used RNA-Seq-based liver transcriptomic profiling of high- and low-RFI dairy cows in two breeds and identified significantly DEGs, their molecular mechanisms, their interactions with other genes and functional enrichments of different molecular pathways. The DEGs that were identified were the *CYP*'s and *GIMAP* genes for the Holstein and Jersey cows, respectively, which are related to the primary immunodeficiency pathway and play a major role in feed utilization and the metabolism of lipids, sugars and proteins.

Keywords: RNA-Seq, Feed efficiency, Residual feed intake, Differentially expressed genes, Pathways, Dairy cattle

* Correspondence: hajak@dtu.dk
[4]Department of Bio and Health Informatics, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark
Full list of author information is available at the end of the article

Background

Feed efficiency is an important trait that should be improved to increase the sustainability and profitability of livestock production. On the one hand, there is a growing demand for food derived from dairy cattle; on the other hand, this production is associated with a high carbon footprint [1, 2], affecting the sustainability of dairy farming. Thus, there is a call for more long-term sustainable interventions. Animal genomics, particularly research regarding the potential genes that are differentially expressed in relation to an increased or a decreased efficiency of feed utilization in dairy cattle, could contribute towards achieving these goals [3]. The definition of feed efficiency in dairy animals is more complicated than that in growing animals because the catabolism of body reserves, followed by the anabolism of body reserves until the next calving period, must be considered in dairy animals [4]. The main purpose of dairy cattle is the production of milk, and it is important to select cattle that have a high efficiency in converting feed into milk. This high efficiency will lead to lower feed costs and increased profits for milk producers [5]. High feed intake and feed efficiency reflect the high production of milk (yield, fat content, protein, lactose and other milk contents) [6]. Therefore, measuring the feed efficiency is important to improve the environment and profits of milk producers.

Feed efficiency is conventionally evaluated using a conversion ratio of the feed intake to the output of the cows. Feed conversion efficiency is an expensive trait to assess and, thus, lends itself to genomic selection. Moreover, it is not sufficient to measure how much nutrients the animal uses to convert into energy to support growth, lactation and body maintenance. In the last 10 years, transcriptomics in dairy cattle has used gene expression microarrays to identify candidate genes for milk yield, protein yield, fertility and metabolic diseases, such as ketosis and milk fever [7–10]. However, only a few studies have focused on liver transcriptomic data of feed efficiency in dairy cattle, and none have focused on Nordic dairy cattle [9–11].

Residual feed intake (RFI) has been used to describe feed efficiency in animals, including beef and dairy cattle [12–15]. Residual feed intake has been defined as the difference between the actual and predicted feed intake [16]. In other words, animals with low RFI are more feed efficient compare to high RFI animals. The heritability of the RFI trait (between 0.01 and 0.38) is quite reliable as a genetic selection trait [17–19]. Hence, the RFI may be a relevant trait to consider in selecting genetically superior animals for breeding studies. Genome-wide association studies (GWAS) characterizing the gene expression and gene regulatory mechanisms related to feed efficiency are quite established in pigs (and poultry) [20, 21]; however, such studies in dairy cattle are fairly recent [22]. In this study, we used an RFI adjusted for stage of lactation, management group, breed and parity. Given the major role of the liver in regulating nutrient homeostasis [23], it is important to understand the biological mechanisms underlying this process. Thus, genome-wide gene expression studies of the liver can provide biological insights into feed processing efficiency and help to determine the mechanism(s) of feed efficiency.

Transcriptomic analyses are useful for studying animal production and health [24] and have become important components of systems genomic or systems biology methods [25]. Transcriptomic analyses provide a snapshot of all the gene expression profiles in a given tissue and insight into the gene functions pertaining to a particular trait [24]. Microarray technologies have been the main platform for animal science research in recent years; however, this trend has been increasingly replaced by RNA-Seq technologies [24–26].

The primary objective of the present study was to identify potential regulatory genes and molecular pathways involved in RFI in dairy cattle by characterizing the liver transcriptome based on RNA-Seq technologies [24, 26]. Another objective of this study was to evaluate the effects of different diets interacting with high- and low-RFI cattle and the resulting impact on the gene expression profiles and associated pathways. This study reports important findings regarding potential regulatory genes and the pathways underlying feed efficiency in dairy cattle using next-generation sequencing or RNA-Seq technology and, most importantly, the nutrigenomics aspects of RFI x Diet interactions.

Results

Mapping statistics summary

The sequencing generated, on average, 57,149,474 raw reads (28,574,737 paired reads) per sample. On average, 91% of the read pairs (26,067,856 read pairs) uniquely mapped to the bovine reference genome from the Ensembl database, release 82. On average, 62% of the read pairs mapped to exonic regions, 20% of the read pairs mapped to intronic regions and almost 18% of the read pairs mapped to intergenic regions (Table 1). After quantifying the expression of the 24,616 genes annotated from the *Bos taurus* reference genome, we excluded a total of 12,591 and 12,711 genes from the remainder of the analyses (because of low expression) of the Holstein and Jersey datasets, respectively. In total, 12,025 genes in the Holstein breed and 11,905 genes in the Jersey breed were used for the subsequent analyses.

Differentially expressed genes (DEGs)

The DEGs identified by DESeq2 are shown in the heat map (Figs. 1 and 2).

Table 1 Summary of the average statistics of the sequence quality and alignment information for the Jersey and Holstein breeds

	Jersey	Holstein
Number of input read pairs	29,428,257	28,221,217
Uniquely mapped read pairs	26,386,656	25,749,055
Mapping rate (%)	91.25	91.24
Exonic	62.15	62.08
Intronic	20.19	20.07
Intergenic	17.66	17.86

The interaction analysis showed low numbers of DEGs in both diet groups (Table 2). From the DESeq2 output, 22 genes and 14 genes in the Holstein and Jersey breeds, respectively, were detected as significant DEGs for the interaction between RFI and diet. No significantly DEGs were identified for the interaction in the Holstein group. However, in the Jersey group, two genes, SEC24 Homologue D (*SEC24D*) and FLT3-Interacting Zinc Finger 1 (*FIZ1*), were differentially expressed in the FE groups, depending on the two diet types (Fig. 3).

Furthermore, 70 Holstein and 19 Jersey DEGs were identified by comparing the RFI status directly without accounting for an interaction (Table 2) (Figs. 1 and 2). Nine genes in the Holstein breed and five genes in the Jersey breed were not annotated. The list of DEGs with their fold changes in the Holstein and Jersey cows is shown in Additional files 1 and 2.

Overrepresented pathways and gene networks

The GOseq analysis did not identify any significantly enriched GO (Gene Ontology) terms or KEGG (Kyoto Encyclopedia Genes and Genomes) pathways.

The output of the GSEA (Gene Set Enrichment Analysis) is presented in Tables 3, 4 and 5, which show the most significantly enriched pathways with FDR (False Discovery Rate) *q*-values less than 0.01. We identified seven overrepresented pathways for the downregulated set of genes, and none were identified for the upregulated genes in the Holsteins. In the Jerseys, two pathways were overrepresented for genes with negative-fold changes, and three pathways were overrepresented for genes with positive-fold changes. The top KEGG pathways for the genes that were downregulated in the high-RFI group in the Holsteins and the Jerseys is the primary immunodeficiency pathway, while the significant pathways identified for the genes that were upregulated in the high-RFI group were only detected in the Jerseys. We also identified that most of the pathways within the strong indication thresholds (FDR *q*-value <0.05) were related to the metabolism of retinols, starch, sucrose, ether lipids and drugs.

The networks identified from the DEGs by IPA® (Ingenuity® Pathway Analysis) are presented in Tables 6 and 7. Seven and six networks were identified for the Holsteins and Jerseys, respectively. The top networks (Fig. 4) in the Holsteins involved 18 genes that are implicated in metabolic diseases, endocrine system disorders and gastrointestinal diseases. The genes that were upregulated in the high-RFI group in the top network of the Holsteins were *ACACA, CYP2C9, CYP7A1, ELOVL6, FOSL2, HCLS1, IFI6, NR1H4, RYR1, SOCS2* and *TBC1D8*; while the downregulated genes were *CR2, CTH, DGAT2, FGFR2, SLC20A1* and *TAF6*.

The top networks in the Jerseys (Fig. 5) involve nine genes that are implicated in cellular compromise, neurological disease, organismal injury and abnormalities. The network includes the genes *CYP3A4, EXTL2* and *TMEM102,* which were upregulated in the high-RFI group, and the genes *FDXR, GIMAP4, GIMAP8, GNG10, HLA*-B and *ZNF613,* which were downregulated in the high-RFI group.

To investigate the DEGs interacting with each other, we analysed the candidate DEGs using the STRING 10 database. Several interacting genes were identified in the Holsteins. In particular, *ACACA* interacts with *BDH2, DGAT2, CYP11A1, HSD17B4, ALDH18A1, HACL1 and ELOVL6*. In the Jerseys, only *GIMAP4* and GIMAP8 interact with each other. The top DEGs present in the IPA network are discussed.

Discussion

Differentially expressed genes

The liver plays an important role in regulating the nutrient supply [27]. Hence, the liver transcriptome may lead to the identification of genes that are important for regulating feed efficiency [28, 29]. Understanding the mechanisms of action and biological functions of the highly significant DEGs in high- versus low-RFI animals experimentally tested under controlled versus high concentrate diets improves our understanding of the biology of feed efficiency in dairy cattle.

The results of this study show a robust relationship and interaction between certain genes involved in feed utilization, partitioning of energy and metabolism. The potential regulatory genes that show a positive effect on RFI were reported in this study.

Almost all the DEGs in the interaction analysis were also present in the analysis without the interaction term. This result may be due to the treatment diet (either low or high concentrate), which might not have a significant impact or be reflected in the differences in the gene expression in the Holsteins. A similar effect was observed in the Jerseys. However, we obtained a smaller number of DEGs compared to those in the Holsteins, which could be due to the small variation among the individuals in the Jersey high and low RFI groups. However, it should be noted that the number of

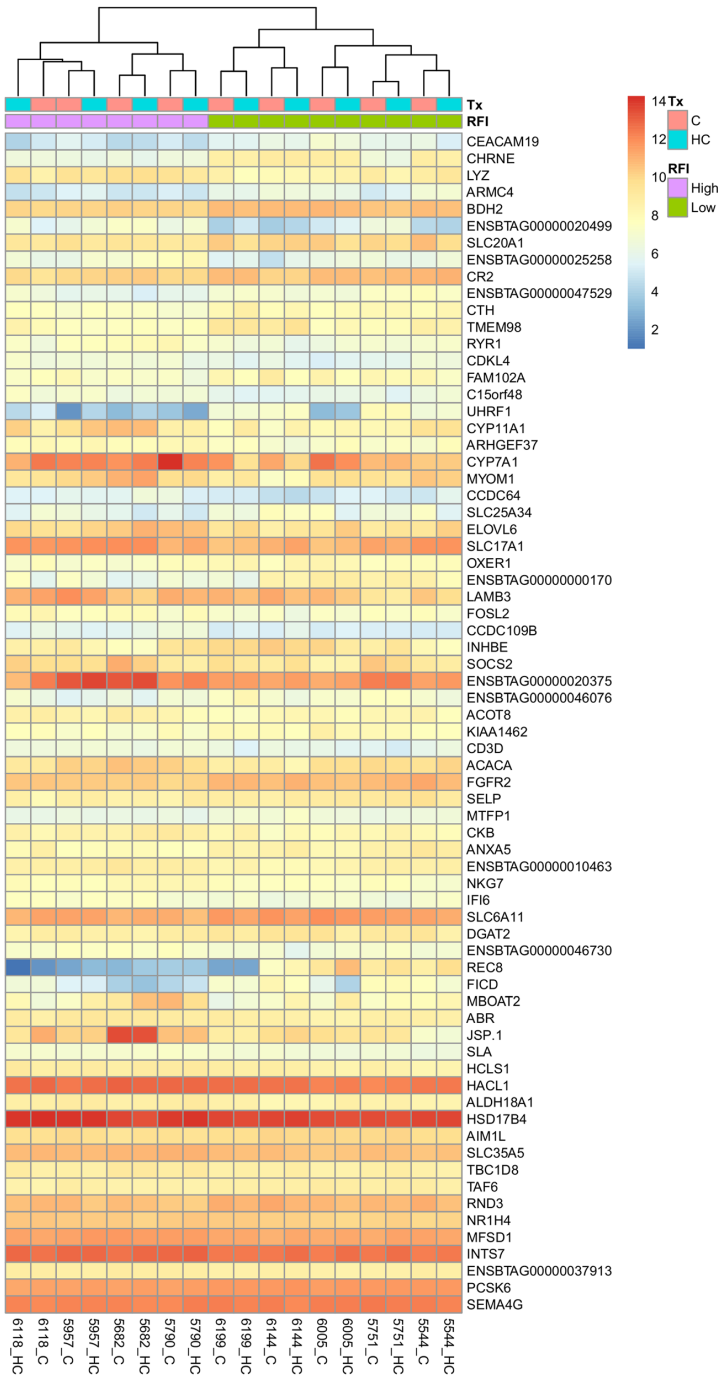


Fig. 1 Heatmap showing the gene expression data of the 70 significantly differentially expressed genes (padj < 0.05) annotated with the gene ID in Holsteins without the interaction term. The data are log2 normalized. Tx = Treatment diet; C = Control diet; HC = High concentrate diet; RFI = Residual Feed Intake groups

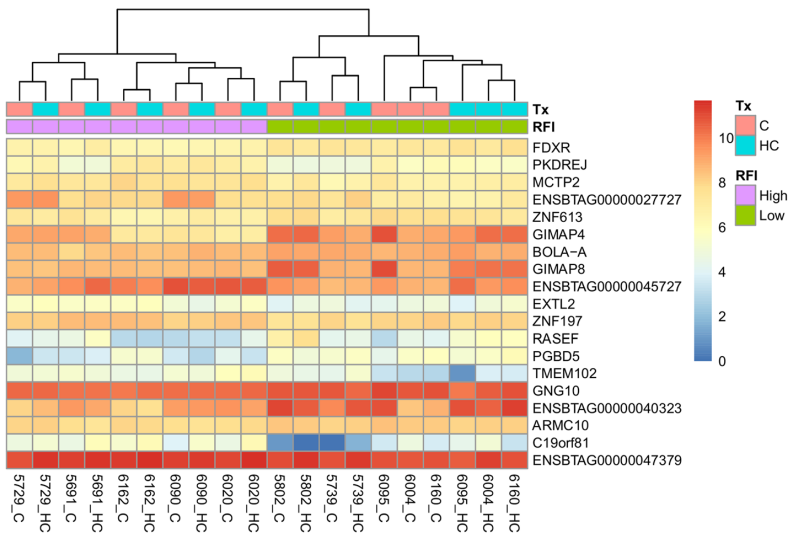


Fig. 2 Heatmap showing the gene expression output of the 19 significantly differentially expressed genes (padj < 0.05) annotated with the gene ID in Jerseys without the interaction term. The data are log2 normalized. Tx = Treatment diet; C = Control diet, HC = High concentrate diet; RFI = Residual Feed Intake groups

animals from each breed is rather small and could have biased the results.

Significantly enriched GO terms and pathways were not identified by GOseq; therefore, we focused on a number of genes that appeared several times in significant networks in the IPA, GSEA and STRING 10. Hence, *ACACA* (Acetyl-CoA Carboxylase Alpha), *CYP11A1* (Cytochrome P450, Family 11, Subfamily A, Polypeptide 1), *CYP2C9* (Cytochrome P450, Family 2, Subfamily C, Polypeptide 9) *BDH2* (3-Hydroxybutyrate Dehydrogenase, Type 2), *DGAT2* (Diacylglycerol O-Acyltransferase 2), and *FBP2* (Fructose-1,6-Bisphosphatase 2) in the Holsteins and *CYP3A4* (Cytochrome P450, Family 3, Subfamily A, Polypeptide 4) and *FDXR* (Ferredoxin Reductase) in the Jerseys were chosen to gain a better understanding of the role of the top genes and networks that were involved. Some of the DEGs reported in previous reports [10, 28, 30] were found to be involved in similar processes related to feed utilization in humans, ruminants and other mammals.

Table 2 Number of differentially expressed genes between high- and low-RFI in a separate diet group in the model with an interaction term and without an interaction term (the diet group was pooled together) according to corrected *p*-values < 0.05

	Control	High concentrate	With interaction	Without interaction
Holstein	9	13	0	70
Jersey	6	6	2	19

Recently, an investigation of two divergent RFI groups in beef cattle using RNA-Sequencing [28, 31] revealed eight and seven significantly DEGs, respectively. However, similar DEGs were not identified in the present study on dairy cattle, suggesting that the discrepancy may be based on the breed. However, some of our results are consistent with a study showing a connection between immune function and most of the DEGs associated with low and high RFI in beef cattle [31]. Alexandre et al. (2015) [28] concluded that the DEGs related to feed efficiency and hepatic physiology were focused more towards the immune response, the metabolism of lipids and cholesterol and hepatic inflammation, which is also consistent with the findings of the present study.

Insights from the Gene Set Enrichment Analysis (GSEA)

Primary immunodeficiency was the top overrepresented pathway detected by the GSEA. This pathway is present and significantly enriched in both cattle breeds. It was stated in the details of the pathway that primary immunodeficiency is a heterogeneous group of disorders. The downregulation of the primary immunodeficiency pathway in the high-RFI cows in both breeds suggests that a low immunity may affect the efficiency of feed utilization. Ozuna et al. (2012) [32] observed that primary immunodeficiency disorder is consistently inherited by low-feed efficiency pigs. Consistently, Kogelman et al. (2014) [33] and Do et al. (2013) [34] reported a correlation between genes related to immunodeficiency function

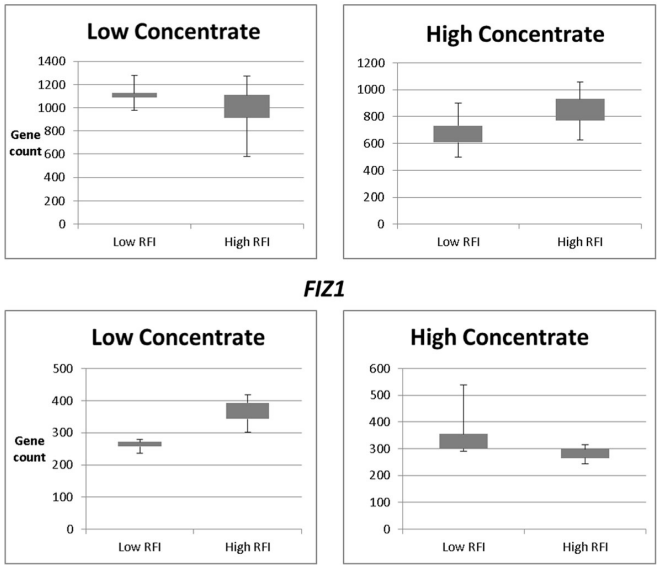


Fig. 3 The plot counts of the 2 genes that show a significant change (padj < 0.05) greater than 0.5-fold in the interaction analysis in the Jerseys

disorders or immunity-related diseases and low-feed efficiency in pigs.

The results of the enrichment and pathway analysis of the DEGs contributes towards the understanding of the function of these genes in relation to the efficiency of feed utilization. The steroid hormone biosynthesis pathway was one of the top KEGG pathways identified in an analysis of negative energy balance in dairy cows [10]. We also discovered that this pathway was overrepresented in the set of genes that were upregulated in the high-RFI group in the Jersey cows (FDR q-value < 0.05). Steroid hormone biosynthesis should always occur in the adrenal

Table 3 KEGG pathways identified for the downregulated genes in the high-RFI group with an FDR q-value < 0.01 from the output of the GSEA in the Holsteins

	Name	FDR q-value	Core enrichment gene
1	Primary immunodeficiency	~0	*CD3D, IL7R, PTPRC, JAK3, ZAP70, CD3E, LCK, ADA, CD8A, BTK, TAP1, UNG, RFX5, CD4*
2	Natural killer cell mediated cytotoxicity	~0	*NFATC2, TNFRSF10D, NCR3, ICAM1, RAC2, ZAP70, PIK3CG, GRB2, LCK, NFAT5, PTK2B, LCP2, PRF1, ITGAL, TYROBP, PIK3CD, SH2D1A, TNF, VAV1, TNFSF10, PLCG2, ITGB2, PAK1, PIK3R5, KRAS, PRKCA, FASLG, SYK, LAT, CD48, IFNGR1, PIK3CA, FCER1G, KLRK1, RAF1, PTPN11, FAS, IFNAR1, PTPN6, HRAS, SOS2, PRKCB*
3	T cell receptor signaling pathway	~0	*CD3D, NFATC2, JUN, PTPRC, ITK, CD3G, ZAP70, CD3E, PRKCQ, PIK3CG, GRB2, LCK, NFAT5, CD8A, RASGRP1, LCP2, TEC, CARD11, PIK3CD, TNF, VAV1, NFKBIA, PAK1, PIK3R5, KRAS, MAPK9, CD4, PDK1*
4	Leukocyte transendothelial migration	0.002	*GNAI1, NCF1, RAPGEF4, OCLN, ITK, CLDN2, ICAM1, RAC2, CLDN1, NCF4, PIK3CG, CDH5, CXCL12, EZR, PTK2B, ITGAM, CLDN4, CYBB, ITGAL, NCF2, PIK3CD, VAV1, CLDN15, PLCG2, ITGB2, PIK3R5, PRKCA, MYL12B, ARHGAP35, F11R, ROCK2, RAP1A, ITGB1, ITGA4, PIK3CA, CXCR4, MSN, CTNNB1*
5	Chemokine signaling pathway	0.002	*CXCR6, GNAI1, CCR2, NCF1, ITK, CXCL9, DOCK2, CCL14, PLCB2, RAC2, JAK3, HCK, PIK3CG, GRB2, CX3CR1, CXCL12, ADCY7, ELMO1, PTK2B, GRK5, CCR5, WAS, ARRB1, PIK3CD, ADRBK2, VAV1, LYN, NFKBIA, PAK1, PIK3R5, KRAS, GNB4, GNG2, PRKX, FGR, STAT3, ROCK2, GNB5, RAP1A, PLCB1, STAT1, IKBKG, AKT3, CHUK, PIK3CA, CXCR4, GNG10, PRKACB*
6	FC gamma R mediated phagocytosis	0.008	*NCF1, SCIN, PTPRC, DOCK2, RAC2, HCK, ARPC1B, MYO10, PIK3CG, MARCKS, LIMK1, PLA2G4A, WAS, INPP5D, PIK3CD, ASAP1, VAV1, PLCG2, LYN, PAK1, PIK3R5, PRKCA, SYK, ARPC1A, PIKFYVE, LAT, PLD2, ARPC3, AKT3, PIK3CA*
7	Propanoate metabolism	0.009	*ACACA, ACSS2, ACAT2, ACACB, EHHADH*

FDR q-value = adjusted p-value; core enrichment gene = subset of genes that contributes most to the enrichment result

Table 4 KEGG pathways identified for the downregulated genes in the high-RFI group with an FDR *q*-value < 0.01 from the output of the GSEA in the Jerseys

	Pathways name	FDR *q*-value	Core enrichment gene
1	Leukocyte transendothelial migration	0.006	*NCF1, PLCG1, ICAM1, VAV1, MAPK12, MSN, PTK2, SIPA1, ITGAM, NCF4, RAP1B, VASP, PIK3CD, RHOH, PIK3R5, RAC2, RAPGEF3, ITK CLDN14, THY1, MYL12B, CLDN4, CXCR4, ACTG1, ITGB2, CYBA, CLDN7, EZR, CYBB, CLDN1, GNAI1, NCF2, MMP2, PRKCB*
2	Primary immunodeficiency	0.010	*CD4, CD8A, ADA, PTPRC, JAK3, TAP1, ZAP70, CD3e, CD3D, LCK*

FDR *q*-value = adjusted *p*-value; core enrichment gene = subset of genes that contributes most to the enrichment result

glands and gonads, while the liver is the site of steroid hormone inactivation. The upregulation of this pathway indicated that steroid hormones were inactivated in the high-RFI group. Therefore, we could conclude that this pathway plays an important role in FE. Furthermore, both *CYP11A1* and *CYP7A1,* which function in cholesterol homeostasis, were identified as DEGs in our experiments, and they are a part of this KEGG pathway.

Additional interesting KEGG pathways that were upregulated in the high-RFI Jersey group were involved in xenobiotics metabolism, retinol metabolism, sphingolipid metabolism, starch and sucrose metabolism, ether lipid metabolism, arachidonic acid metabolism and drug metabolism cytochrome P450. Most of these pathways (Additional file 3) were related to nutrients (fatty acids, carbohydrates and proteins) and metabolism. de Almeida Santana et al. (2016) [35] reported that the retinol metabolic pathway was involved in the feed conversion ratio in beef cattle in relation to rump fat thickness. The authors also discussed that lipid and protein metabolisms were well-known important factors in feed efficiency physiology. The relationship between retinol metabolism and the feed conversion ratio phenotype in Nellore beef cattle has been previously described [36] and [35].

The top pathway of the metabolism of xenobiotics by cytochrome P450 involved the *CYP* genes. Specifically, the *CYP11A1* gene was upregulated in the high-RFI group compared with that in the low-RFI group. The *CYP11A1* gene was not present in the IPA output because it has no Entrez gene ID when uploaded as an input. However, *CYP11A1* was also identified as a DEG in the Holstein group. The *CYP11A1* gene is also known as cytochrome P450, which functions in drug metabolism and cholesterol, steroid and lipid synthesis. When the expression of this gene is high, it will also lead to the active synthesis of lipids, steroids and hormones. Yi et al. (2015) [29] have mentioned that the upregulation of *RSAD,* which is a gene that has a similar function to *CYP11A1* in the low RFI (high feed efficiency) group, may lead to a decreased feed intake, high energy utilization and few energy costs by modulating fatty acid and leptin metabolism. These results are consistent with those reported by McCabe et al. (2012) [10], who discovered that *CYP11A1* was upregulated in severe negative energy balanced cows. This result suggests that the *CYP11A1* gene indeed played an important role in lipid synthesis and the regulation of cholesterol synthesis in the liver. Together with *CYP11A1*, the *CYP7A1* and *CYP2C9* genes were also differentially expressed and had the same pattern of expression in the Holsteins. In another study conducted by [37], the *CYP* genes were involved in steroidogenesis and converted cholesterol into pregnenolone and then to dehydroepiandrosterone (DHEA). The *CYP* gene function was also discussed in feed efficiency, particularly pertaining to hepatic metabolism [28, 38, 39].

Ingenuity® Pathways Analysis (IPA®) output and interactions between DEGs

The output of the IPA for the Holsteins showed the top networks of the 18 upregulated DEGs, which included Metabolic Diseases, Endocrine System Disorders and Gastrointestinal Diseases. Consistently, the network of metabolic diseases was associated with the differential gene expression in the severe negative energy balance in high-yielding cows [11]. The metabolic disease network

Table 5 KEGG pathways identified for the upregulated genes in the high-RFI group with an FDR *q*-value < 0.01 from the output of the GSEA in the Jerseys

	Pathways name	FDR *q*-value	Core enrichment gene
1	Retinol metabolism	0.002	*PNPLA4, CYP2B6, CYP2C18, RETSAT, CYP1A1, RDH11, CYP1A2, ALDH1A1, CYP26B1, LRAT, ADH5, UGT2A3, RDH16, UGT1A1, ALDH1A2, RDH10*
2	Metabolism of xenobiotics by cytochrome P450	0.003	*CYP2B6, ALDH1A3, CYP2E1, CYP2C18, EPHX1, CYP1A1, CYP1A2, MGST1, MGST3, ALDH3B1, ADH5, UGT2A3, UGT1A1*
3	Ether lipid metabolism	0.009	*ENPP6, PLA2G7, PLD2ENPP2, LPCAT2PLA2G12A, AGPSPLD1*

FDR *q*-value = adjusted *p*-value; core enrichment gene = subset of genes that contributes most to the enrichment result

Table 6 Gene networks from the 58 differentially expressed genes for the Holstein group converted to human orthologous genes

ID	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
1	**ACACA**, *Akt* ,**CR2**, **CTH**, **CYP2C9**, **CYP7A1**, **DGAT2**, **ELOVL6**, **FGFR2**, **FOSL2**, Growth hormone, **HCLS1**, **IFI6**, *IFN* Beta, *IL1*, Immunoglobulin, Insulin, Interferon alpha, *LDL*, **LYZ**, N-cor, Nr1h, **NR1H4**, *PDGF BB*, *Pkc*(s), RNA polymerase II, *Rxr*, **RYR1**, **SLC20A1**, **SOCS2**, *STAT5a/b*, **TAF6**, **TBC1D8**, *TCR*, *Vegf*	41	18	Metabolic Disease, Endocrine System Disorders, Gastrointestinal Disease
2	*ACOT2*, **ACOT8**, **AIM1L**, **ALDH18A1**, *AQP8*, **C15orf48**, **CCDC64**, cholesterol, *E2F3*, **FAM102A**, *GSTO1*, *HIST3H2A*, *HNF4A*, **HSD17B4**, *HSPB1*, *INHBC*, **INHBE**, **KIAA1462**, *LPCAT3*, **MFSD1**, *MYC*, *NBEA*, palmitoyl-CoA hydrolase, **PCSK6**, **REC8**, *RNF2*, *SLC25A13*, **SLC35A5**, *SMC1B*, *STAG3*, *SYCP1*, *SYCP3*, *TGFB1*, **UHRF1**, **Wfdc21**	32	15	Cellular Assembly and Organization, DNA Replication, Recombination, and Repair, Cell Cycle
3	**ABR**, **ANXA5**, **BDH2**, caspase, *CD3*, **CD3D**, *CHRNA1*, *CHRNA2*, *CHRNA3*, *CHRNB1*, *CHRNB4*, *CHRND*, **CHRNE**, *CHRNG*, *DDT*, *DLG4*, *ERK*, *GABPA*, *HIST1H2AC*, *HS2ST1*, *Il8r*, **INTS7**, *KRAS*, nicotinic acetylcholine receptor, *NRG1*, *NRG* (family), *P38 MAPK*, *PI3K* (complex), **RND3**, **SEMA4G**, **SLA**, **TMEM98**, *TRD*, *TRG*, *VAV1*	20	10	Haematological Disease, Metabolic Disease, Organismal Injury and Abnormalities
4	*CASP1*, **CCDC109B**, **CDKL4**, *COL7A1*, *Cr3*, *CSTB*, *CYCS*, *DUSP3*, *ERK1/2*, **FICD**, *GNAI2*, **HACL1**, *HLA* Class I, Hla-abc, **HLA-B**, *HOMER3*, *HSP90AA1*, *HTT*, *IARS2*, *IL22RA2*, *LAMA3*, **LAMB3**, *MAGEB6*, *MAST2*, **MBOAT2**, *MCU*, **MTFP1**, **SLC17A1**, **SLC6A11**, *SMAD3*, tauroursodeoxycholic acid, *TEAD2*, *TPP2*, *WISP1*, *ZFYVE27*	19	10	Connective Tissue Disorders, Dermatological Diseases and Conditions, Developmental Disorder
5	**ARHGEF37**, *BIN3*	2	1	Cell Morphology, Embryonic Development, Organ Development
6	5(S)-hydroperoxyeicosatetraenoic acid, **OXER1**	2	1	Cell Morphology, Cellular Development, Tissue Development

Red= Upregulated in high-RFI; **Green**=Downregulated in high-RFI

Table 7 Gene networks from the 15 differentially expressed genes for the Jersey group converted to human orthologous genes

ID	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
1	*ALDH2, APP, ATP6V1A, C1orf112, CYB5B,* ***CYP3A4****, DCTPP1, DDT, DUSP3 ,ETFB,* ***EXTL2****,* ***FDXR****,* ***GIMAP4****,* ***GIMAP8****, GJC1, GLRX5, GNB1,* ***GNG10****, HLA* Class I, Hla-abc,* ***HLA-B****, HSPA1A/HSPA1B, HSPE1*, hydrogen peroxide, *IARS2, MAPK1, NFkB* (complex), *PSEN1, STAT6,* ***TMEM102****, TRAPPC4, TRAPPC9, TRAPPC6B, TXNDC17,* ***ZNF613***	23	9	Cellular Compromise, Neurological Disease, Organismal Injury and Abnormalities
2	*GATA2,* ***MCTP2****, Pdzph1*	6	2	Cancer, Cell-mediated Immune Response, Cellular Development
3	***ARMC10****, MME*	3	1	Cancer, Cell Death and Survival, Organ Morphology
4	*RABL2B,* ***RASEF***	3	1	Cancer, Haematological Disease, Immunological Disease
5	*LNX1, LNX2,* ***PKDREJ***	3	1	Protein Synthesis, Haematological System Development and Function, Humoral Immune Response
6	*FBXL19,* ***PGBD5****, RABGGTB*	3	1	Post-Translational Modification, Cellular Movement, Respiratory System Development and Function

Red= Upregulated in high-RFI; **Green**=Downregulated in high-RFI

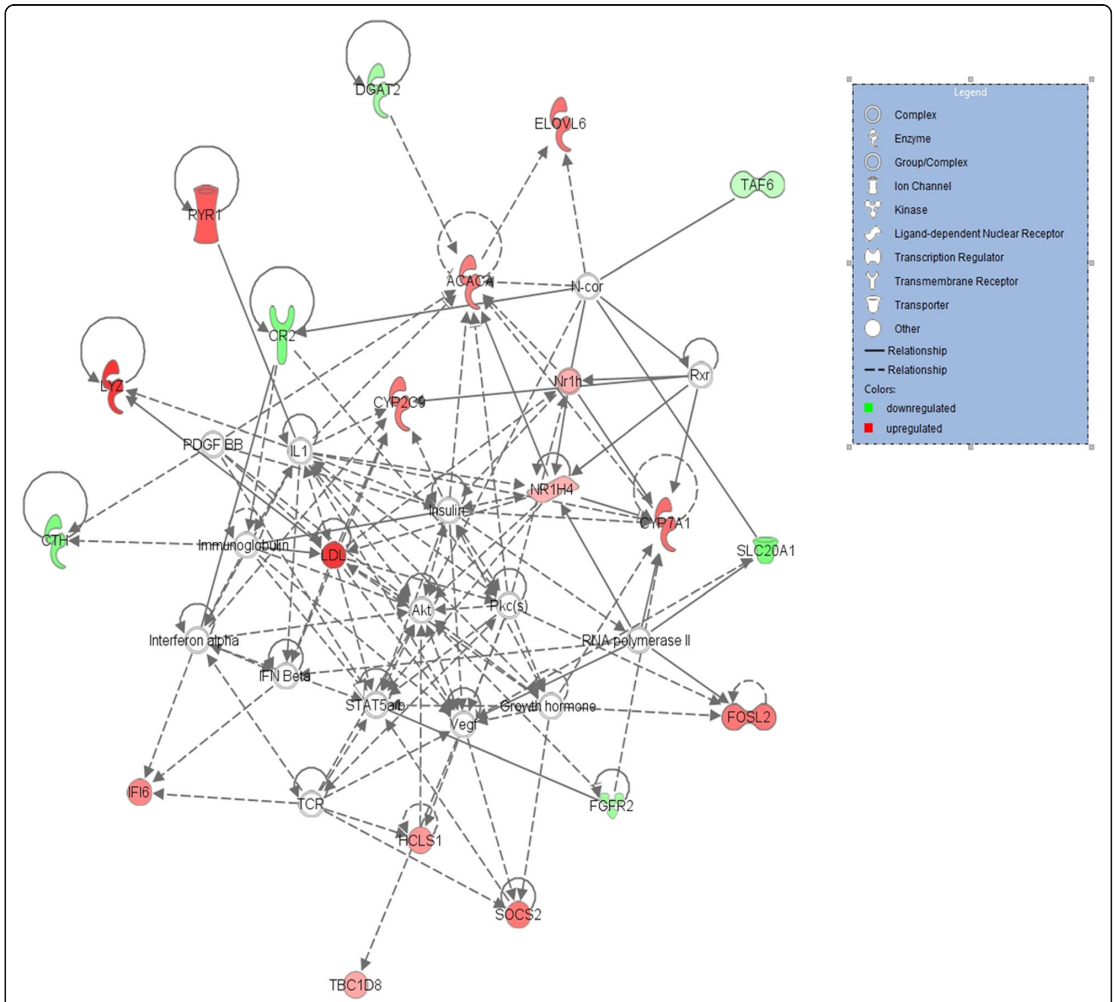


Fig. 4 The relationship between 18 DEGs in network 1 in the Holsteins. The genes highlighted in red were upregulated, while those highlighted in green were downregulated in the high-RFI group

may be closely related to the immune system. Paradis et al. (2015) [31] have stated that immunity is very important to produce animals that have less energy to fight against systemic inflammation, have better detoxification of endotoxins and use more energy for growth.

The output from the STRING 10 analysis shows that among the significantly DEGs, the *ACACA* gene has interactions with *CYP11A1, BDH2, DGAT2, HSD17B4, FGFR2, HACL1* and *ALDH18A1*. This interaction depicts the importance of the *ACACA* gene in this output. The function of the *ACACA* gene is to convert acetyl CoA to fatty acids, also known as lipogenesis. The upregulation of the *ACACA* gene in the high-RFI Holstein group in this network is also interesting in relation to functions in feed utilization. A positive relationship has been reported between ACACA enzyme activity and intramuscular fat levels [40]. In addition, the negative relationship between the *ACACA* gene and other lipogenesis pathway genes and milk production in dairy cattle was also confirmed by Sumner-Thomson et al. (2011) [41]. Hence, the increased *ACACA* gene expression might reflect the deposition of fat in the high-RFI cows.

The output of the DEG analysis revealed that *BDH2* is another interesting gene to be considered due to its downregulation in the high-RFI cattle. These genes play an important role in metabolism and synthesis and are very well known for their role in the degradation of ketone bodies. In contrast, no change was observed in the

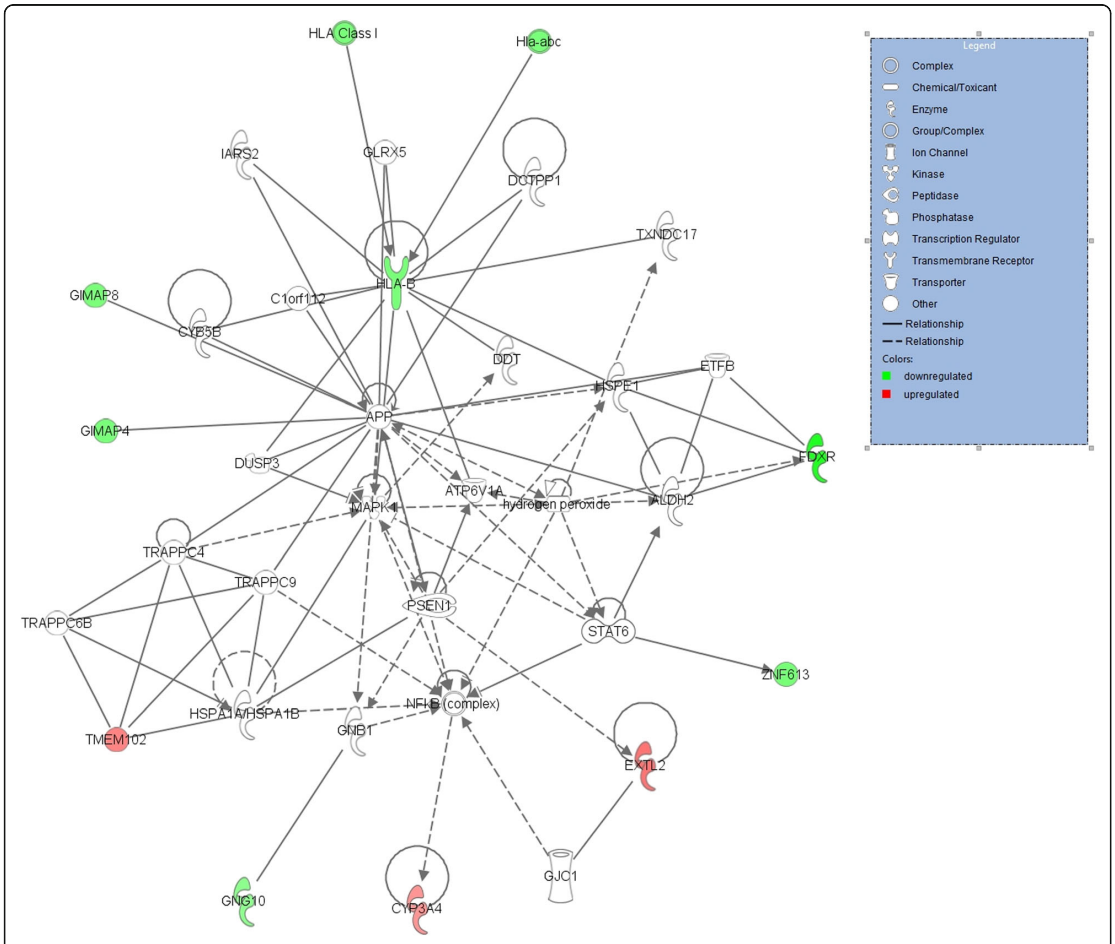


Fig. 5 The relationship between the nine DEGs in network 1 in the Jerseys. The genes highlighted in red were upregulated, while those highlighted in green were downregulated in the high-RFI group

transcript abundance of genes involved in ketone body synthesis [3-hydroxy-3-methylglutaryl-CoA synthase 2 (*HMGCS2*), 3-hydroxybutyrate dehydrogenase, type 2 (*BDH2*)] in cows subjected to nutrient restrictions to reduce the frequency of milking [42]. However, our results showed a downregulation of *BDH2* genes in the high-RFI cattle, suggesting that this group was inefficient in degrading ketone bodies.

In the present study, an upregulation of the *DGAT2* gene was observed in the high-RFI Holsteins. In humans, the *DGAT2* gene was reported to be a candidate for the dissociation between fatty liver and insulin resistance [30], and this result has also been observed in mice [43]. The *DGAT* gene functions in the liver by catalysing the final reaction in the synthesis of triglycerides in which diacylglycerol is covalently bound to long chain fatty acyl-CoAs. The *DGAT* gene might be a candidate for treating obesity in humans because the increased expression of *DGAT* led to obesity in mice that were resistant to diet-induced obesity [44].

The IPA analysis output for the Jersey breed showed the top overrepresented networks, involving nine DEGs that are related to cellular compromises, neurological disease, organismal injury and abnormalities. These processes appear to be closely related to the top primary immunodeficiency output from the GSEA KEGG pathways. The importance of this related pathway was previously explained as it pertains to Holsteins. The output from STRING 10 showed only one interaction between the *GIMAP4* and *GIMAP8* genes among the significantly

DEGs in the Jerseys. Although the Jersey DEGs differed from those in the Holsteins, some genes have similar functions, such as the *CYP3A4* gene.

The ferredoxin reductase (*FDXR*) gene was also a top DEG in the Jersey breed. The *FDXR* gene encodes a 50,000 kDa mitochondrial flavoprotein attached to the matrix side of the inner mitochondrial membrane. FDXR transports electrons from NADPH via the soluble single electron shuttle ferredoxin to a membrane-integrated cytochrome P450 enzyme (*CYP11A1*). The upregulation of the *FDXR* gene, which occurs in the low-RFI Jersey group, can deplete the levels of the reduced NADPH. This *FDXR* gene is also known to be involved in cholesterol metabolism, which is also a part of steroid metabolism.

In addition, it is interesting to note that the *GIMAP4* and *GIMAP8* genes, which were upregulated in the low-RFI Jersey group, are also related to an immunoassociated nucleotide (IAN) subfamily of nucleotide-binding proteins. This is important for controlling the immune system and responding to infections [45]. These genes have never been implicated or previously described in relation to feed efficiency or utilization in any species. The expression consistency of these two genes is interesting to relate to the biological functions that are important for controlling the immune system. Consistent with the GSEA output results, primary immunodeficiency is the top pathway and is reflected by the differential expression of these two related genes. The *GIMAP4* and *GIMAP8* genes require further investigation regarding their importance in controlling the immune system.

Genes in the RFI x Diet interaction in the Jersey cattle

The DEGs involved in the interaction between RFI and diet were also associated with immunodeficiency, which was a key pathway consistently identified in this study. It is interesting that the diet has an impact on genes belonging to the immunodeficiency pathway, and this result paves the way for future studies to determine how to improve diet in relation to the genetic background of the animals. Two protein-coding genes, *SEC24D* and *FIZ1,* were differentially expressed in response to the diet and were associated with pathways, including Immune System and Transport to the Golgi and subsequent modification and were involved in transcriptional regulation [45]. These genes might also be factors in the primary immunodeficiency pathway that was detected as significantly overrepresented in this study. The lack of a more extensive differential gene expression response indicates that the differences in the concentrate composition of the diet tested in this analysis may not have been sufficient to influence gene expression levels.

Implications for improving feed efficiency via breeding

Through the integration of the information obtained from the DEGs, functional enrichment, pathway analysis and published data, this study provides a list of candidate genes whose functions and expression levels are strongly related to RFI. These candidate genes can be used to develop genomic biomarkers, eQTLs (expression quantitative trait loci), CNV (Copy Number Variation), SNPs (single-nucleotide polymorphisms) and additional markers for possible inclusion in genomic selection methods utilizing functional information [e.g., sgBLUP (system genomic BLUP) [24] and BLUP|GA (BLUP approach given the Genetic Architecture] [26].

This study was conducted with relatively small sample sizes (10 samples in each breed) but in a highly controlled environment. However, it is recommended that this study should be replicated with a larger sample size for the eventual validation of our findings.

Conclusion

This study investigated the liver transcriptome of high- versus low-RFI animals experimentally tested with control versus high concentrate diets. The results provide an important understanding of the biology of feed efficiency in dairy cattle and a basis for elucidating the mechanisms of action and biological functions of highly differentially expressed genes. This study is novel in at least two aspects as follows: one in terms of the species/breed (dairy cattle: Danish Holsteins and Danish Jerseys) and the second in terms of the RFI x Diet experiments. Furthermore, to the best of our knowledge, this study is the first study conducted exploring residual feed intake in Nordic dairy cattle using RNA-Seq, which is known as the most accurate technology for genome-wide gene expression studies. The results reveal differences in the biological mechanisms related to residual feed intake in the Holsteins and Jerseys. The study identified 70 and 19 candidate genes that are involved in the regulation of feed efficiency pathways in the Holstein and Jersey cattle, respectively. The candidate genes identified in this study will be useful for explaining the biological effects of genomic markers in genomic selection methods utilizing functional information.

Methods

Animal ethics statement

In this study, individual cows of the two main dairy cattle breeds in Denmark, Holstein and Jersey, were obtained from Danish Cattle Research Centre (DCRC), Aarhus University, Denmark. The data from this herd have previously been used in quantitative genetic studies regarding feed or dry matter intake [46]. The experimental animal procedures were approved by the Danish Animal Experimentation Inspectorate.

Table 8 Details of the experimental cows. The cows have been classified according to the breed, parity, block, RFI value, RFI group and the allocation of the diet for the first and second period. RFI values refer to the random animal solutions as explained in the text

Cow ID	Breed	Parity	Block	RFI value	RFI group	1st period	2nd period
6199	Holstein	1	5	-0.395	High	HC	C
5751	Holstein	3	2	-0.622	High	HC	C
6118	Holstein	1	3	-0.03	Low	HC	C
5957	Holstein	2	1	0.885	Low	HC	C
5790	Holstein	2	4	0.101	Low	HC	C
6004	Jersey	2	1	-1.705	High	HC	C
5739	Jersey	3	4	-0.042	High	HC	C
6090	Jersey	1	5	0.493	Low	HC	C
6162	Jersey	1	3	0.803	Low	HC	C
5729	Jersey	3	2	0.938	Low	HC	C
6144	Holstein	1	3	-1.103	High	C	HC
6005	Holstein	2	1	-1.046	High	C	HC
5544	Holstein	3	4	0.05	High	C	HC
5682	Holstein	3	2	0.695	Low	C	HC
6160	Jersey	1	5	-0.511	High	C	HC
6095	Jersey	1	3	-0.401	High	C	HC
5802	Jersey	3	2	-1.048	High	C	HC
6020	Jersey	2	1	0.458	Low	C	HC
5691	Jersey	3	4	2.226	Low	C	HC

RFI = Residual feed intake; *HC* = High concentrate; *C* = Low concentrate (control)

Table 9 Ration composition of the experimental diet

Item	Low Concentrate	High Concentrate
Forage:Concentrate	68:32	39:61
Grass/clover silage (g/kg DM)	684	391
Barley (g/kg DM)	189	377
Rapeseed cake (g/kg DM)	25.7	51.4
Soybean meal (g/kg DM)	85.7	171
Urea (g/kg DM)	4.7	2.7
Mineral premix (g/kg DM)	9.3	5.3
Vitamin premix(g/kg DM)	2.1	1.2
Gross energy (MJ/kg DM)	18.7	19.2
DM (g/kg)	513	620
Ash (g/kg DM)	72.0	57.3
Crude protein (g/kg DM)	170	204
Crude fat (g/kg DM)	31.8	33.6
Starch (g/kg DM)	105	218
Crude fiber (g/kg DM)	179	127
NDF (g/kg DM)	335	271
iNDF (g/kg DM)	45.3	41.8

DM = Dry Matter; *NDF* = Neutral Detergent Fiber; *iNDF* = indigestible Neutral Detergent Fiber

Animals experiments

Ten Jersey and ten Holstein cows were selected from a research herd of 200 animals. However, one of the Holstein cows was excluded from the study due to an unsuccessful liver biopsy. Animals of both breeds were divided into the following two groups: high- or low-residual feed intake (RFI). Residual Feed Intake was defined using the one-step approach [14]. Here, the random animal solutions were extracted from a random regression model in which the dry matter intake was regressed to the following fixed effects: weeks of lactation, the management group in which the cows were held, and the interaction between weeks of lactation, breed and parity. Fixed linear regressions were applied to adjust for the metabolic body weight, daily live weight change, daily body condition score change (fitted with a Legendre polynomial), and energy corrected milk yield. The random effects were cow within the breed and cow within the breed and parity. Cows were ranked based on their random effect solutions. From the available cows, blocks were defined to include two Holstein and two Jersey cows in a similar lactation stage and a similar parity group (first or older), and included one high and one low ranked cow of each breed. In total, five blocks were defined, and the cows within blocks were then allocated to the experimental treatments and measurements.

Table 8 shows the RFI values of the individual cows that were used for the samples and analysis. Table 8 also shows the assignment of the treatments for the first and the second periods of the experiment.

All cows received a low-concentrate [control (C)] and a high-concentrate (HC) diet in a crossover design with two periods (Table 9). There was approximately a 30% difference in the concentrate proportion of the dry matter (DM) basis between the high- and low-concentrate diets. In period 1, five Jersey and five Holstein cows were allocated to the high-concentrate diet, and the other four Holstein and five Jersey cows were allocated to the low-concentrate diet. Then, the animals were placed in four individual open circuit respiration chambers to measure gas exchange during the last 3 days of the trial. However, the measurements of the gases are not presented in this study. On the last day of the diet trial, the cows were transferred to a tie-a-stall area to undergo the liver biopsies.

After the liver biopsies, the cows were transferred and subjected to a new diet. The adaptation to the diets required 14–26 days in period 1 and 14 days in period 2. After the second diet period, another liver biopsy was performed. For the second trial, the cows were placed in a respiration chamber for 2 days at the end of the feeding trial before the transfer for the liver biopsy.

Liver biopsy collection

Ten millilitres (ml) of Procamidor®vet (20 mg/ml) anaesthesia were injected under the skin and into the intercostal muscles at the site of the insertion of the biopsy instrument. Fifteen to 30 min after the injection, the surrounding muscle was numb, and a small incision was made through the skin in preparation for the insertion of the biopsy needle (PRO. MAG™ BIOPSY NEEDLE). Approximately 10–20 mg of liver tissue were collected from the biopsies and immersed in an RNAlater® (Sigma-Aldrich) solution for 6 days and stored at 4 °C. After 6 days, the RNAlater solution was removed, and the tissues were stored at −80 °C until further use.

mRNA extraction and sequencing

mRNA was extracted from the liver tissue samples using the Qiazol, RNeasy® Mini Kit and MaXtract High Density for further RNA-Sequencing.

The quantity and quality of the extracted mRNA were assessed using a NanoDrop® ND-1000 spectrophotometer and Agilent 2100 Bioanalyzer machine. The quantity of the mRNA ranged from 77.95 to 1104.11 ng/μl. The quality of all mRNA samples was above 8 RIN (RNA Integrity Number). The preparation of the cDNA library and the RNA sequencing was performed by AROS Biotechnology A/S (Denmark). The cDNA originating

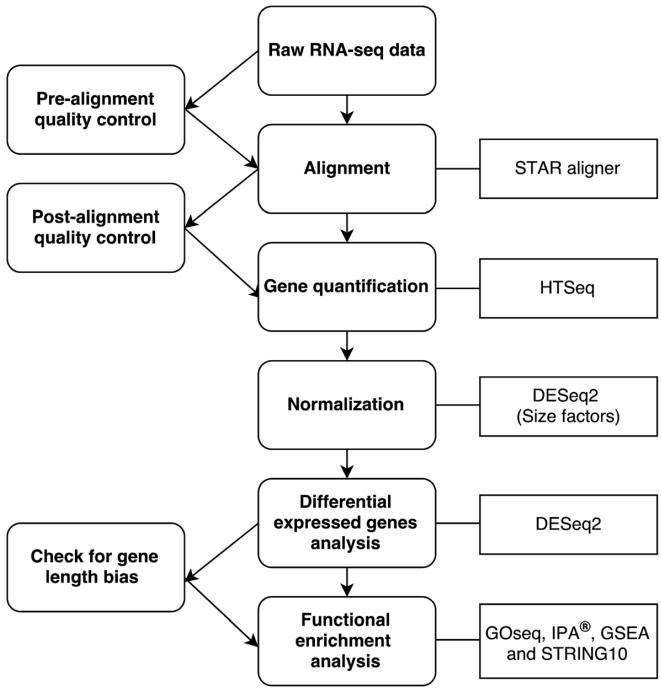


Fig. 6 Working pipeline of the RNA-Seq analysis

from the RNA fragments were paired and sequenced using an Illumina HiSeq 2500 machine, and, on average, 57 million reads per sample were obtained. In detail, the fragments were paired-end sequenced, generating read pairs of 100 bp length and obtaining, on average, 28 million read pairs per sample. The RNA-Seq was performed in one run. All samples (38 samples) were pooled together using four lanes of a flow cell. The raw reads generated from the sequencing machine often were obtained in (or can be converted into) a file format called FASTQ. A read pair denotes that the sequencing was conducted from both ends of the fragment, resulting in a pair of reads, one from each end of the fragment.

Bioinformatics and statistical analysis

The bioinformatics pipeline is shown in Fig. 6. A read quality control was conducted using FastQC version 0.11.3 [47]. Adapters were removed using cutadapt v.1.6f [48], and based on the quality control report, the reads were not further pre-processed.

Reads were aligned to the genome assembly *Bos taurus* UMD3.1 using STAR: ultrafast universal RNA-Seq aligner STAR_2.3.0 [49], providing the *Bos taurus* gene annotation file as additional information. A maximum of five mismatches was allowed, and all the other options were set as STAR default values. The reference genome and annotation file were downloaded from the Ensembl database, release 82.

A post-alignment quality control was performed on the alignment files using Qualimap version 2.0 [50]. The gene expression counts were computed using HTSeq-count [51]. This tool counts the read pairs mapping to a specific gene locus annotated in the Ensembl reference genome. Thus, we generated a matrix for each annotated gene with the corresponding raw counts. We filtered the low count genes, excluding genes with less than 1 count per million (cpm) in at least eight samples for the Holstein group and 10 samples in the Jersey group [31], where eight and 10 were the dimensions of the smallest classes in the treatment control variable in each breed.

Differentially expressed genes (DEGs) were identified using DESeq2 package version 1.12.0 [52].

The gene counts were normalized using the default normalization procedures provided by DESeq2.

The DE analyses were performed separately for each breed. All the parameters were set to default values and fitted with two different models.

$$Model\ 1 \quad Y = Parity\ number + Diet + RFI \qquad (1)$$

where Y is the gene expression counts, RFI is a dummy variable that represents the feed efficiency of the animals, and Parity number and Diet were codified as dummy variables included to control for potentially confounding effects. In this model, we assumed an additive effect without an interaction between diet and RFI.

$$Model\ 2 \quad Y = Parity\ number + Diet + RFI + Diet : RFI \qquad (2)$$

where Y is the gene expression counts, RFI is a dummy variable that represents the feed efficiency of the animals, and Parity number was included as a dummy variable to control for potentially confounding effects. In this model, we assumed an interaction between diet and RFI, and Diet: RFI is the interaction term (2 RFI groups × 2 treatment diets).

Differentially expressed genes were considered at a False Discovery Rate (FDR) < 5%.

A principal component analysis (PCA) was performed using the function plotPCA in the DESeq2 R package to determine the interrelations between the individual samples using the normalized counts of all the genes after filtering as the input. The PCA plot shows a strong effect of the Parity Number. Therefore, the Parity Number was included in the DE analysis to remove its confounding effect.

Functional enrichment analysis

The functional enrichment analysis of the DEGs was performed using the GOseq version 1.24.0 package [53] in R software. Both Gene Ontology (GO) terms and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment were used to find significant enrichment in each DEG set identified. Because of limited annotations of the bovine reference genome, orthologous human genes (Ensembl genes 82) were also used to identify the enriched GO terms and KEGG pathways.

All the significantly DEGs obtained with the DESeq2 package were used as an input for the functional enrichment analysis by QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, http://www.qiagen.com/ingenuity). The Entrez gene ID of a particular gene was used as input. The IPA automatically converts the *Bos taurus* Entrez ID into the corresponding human orthologous gene. We selected the top networks in each species of the network analysis in the IPA.

Finally, an additional analysis was performed using a Gene Set Enrichment Analysis (GSEA) [54, 55] from the Broad Institute that, in contrast to IPA and GOseq, considers the changes in the entire gene profile. It has been previously demonstrated that GSEA provides insight into the biology behind a set of genes in terms of how the DEGs interact with one another [56].

Furthermore, STRING 10 version 10.0 [57] was used to identify interesting associations between the significant genes identified in our study. Using the STRING

database (http://string-db.org/), multiple proteins were chosen from the website interface. The DEG names were inserted as the input in the list of names, and *Bos taurus* was chosen as the organism.

Additional files

Additional file 1: Differentially expressed gene list in Holsteins. (DOCX 23 kb)
Additional file 2: Differentially expressed gene list in Jerseys. (DOCX 17 kb)
Additional file 3: Gene Set Enrichment Analysis output. (DOCX 18 kb)

Abbreviations

BLUP|GA: Best Linear Unbiased Prediction (approach given the Genetic Architecture); C: Control; cDNA: Complementary DNA; CNV: Copy Number Variation; cpm: Count per million; DEG: Differentially Expressed Gene; DM: Dry Matter; eQTL: Expression Quantitative Trait Loci; FDR: False Discovery Rate; FE: Feed Efficiency; GO: Gene Ontology; GSEA: Gene Set Enrichment Analysis; GWAS: Genome Wide Association Study; HC: High Concentrate; iNDF: Indigestible Neutral Detergent Fibre; IPA®: Ingenuity® Pathway Analysis; KEGG: Kyoto Encyclopedia of Genes and Genomes; mRNA: Messenger RNA; NDF: Neutral Detergent Fiber; padj: Adjusted *p* value; PCA: Principal Component Analysis; RFI: Residual Feed Intake; RNA-Seq: RNA-Sequencing; sgBLUP: System genomic Best Linear Unbiased Prediction; SNP: Single Nucleotide Polymorphism; Tx: Treatment

Acknowledgements

The authors thank field staff and laboratory technicians who helped carry out this study.

Funding

The research conducted in this study was funded by project "Feed Utilization in Nordic Cattle" from Danish Milk Levy Foundation, Skejby Denmark. SMS received a stipend for her PhD studies from the Ministry of Education of Malaysia and Universiti Putra Malaysia and her tuition fee waiver from the University of Copenhagen.

Availability of data and materials

The data discussed in this publication were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE92398 at: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92398.

Authors' contributions

HNK was the overall project leader who conceived and conducted this selective transcriptomic profiling study on cows with high/low RFI, and supervised SMS in the laboratory work and bioinformatics/systems biology analyses. PLund and PL supervised the feeding trials and biopsy experiments with the assistance of JKH and DWO. PL provided RFI measurements of cattle in the experiment. SMS processed liver tissue for the RNA isolation as well as quality control of RNA samples before RNA Sequencing. SMS analysed all the data and wrote the first draft of the manuscript with assistance from GM. GM contributed substantially to bioinformatics analyses. All authors wrote, read, and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent to publication

All the authors hereby consent to publish this work in BMC Genomics.

Ethics approval and consent to participate

The experimental animal procedures were approved by the Danish Animal Experimentation Inspectorate.

Publisher's Note

Author details

[1]Animal Breeding, Quantitative Genetics and Systems Biology Group, Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, DK-1870 Frederiksberg C, Denmark. [2]Department of Molecular Biology and Genetics - Center for Quantitative Genetics and Genomics, Aarhus University, AU Foulum, DK-8830, Tjele, Denmark. [3]Department of Animal Science - Animal Nutrition and Physiology, Aarhus University, AU Foulum, DK-8830, Tjele, Denmark. [4]Department of Bio and Health Informatics, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark.

Received: 29 September 2016 Accepted: 11 March 2017
Published online: 24 March 2017

References

1. Rotz C, Montes F, Chianese D. The carbon footprint of dairy production systems through partial life cycle assessment. J Dairy Sci. 2010;93(3):1266–82.
2. Assessment ALC: Greenhouse Gas Emissions from the Dairy Sector.
3. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
4. Herd R, Arthur P. Physiological basis for residual feed intake. J Anim Sci. 2009;87(14_suppl):E64–71.
5. Varga G, Dechow C, Eastridge M. Can We Use Residual Feed Intake to Enhance Dairy Production Efficiency? In: Proceedings of the 22nd Tri-State Dairy Nutrition Conference, Fort Wayne, Indiana, USA, 23-24 April 2013: 2013. Michigan State University; 2013. p. 131-140.
6. Hutjens M. Feed efficiency and its impact on feed intake. eXtension Dairy. 2012.
7. Bionaz M. Nutrigenomics Approaches to Fine-Tune Metabolism and Milk Production: Is This the Future of Ruminant Nutrition? Adv Dairy Res. 2014;24:1–2.
8. Heyen D, Weller J, Ron M, Band M, Beever J, Feldmesser E, Da Y, Wiggans G, VanRaden P, Lewin H. A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol Genomics. 1999;1(3):165–75.
9. Loor JJ, Everts RE, Bionaz M, Dann HM, Morin DE, Oliveira R, Rodriguez-Zas SL, Drackley JK, Lewin HA. Nutrition-induced ketosis alters metabolic and signaling gene networks in liver of periparturient dairy cows. Physiol Genomics. 2007;32(1):105–16.
10. McCabe M, Waters S, Morris D, Kenny D, Lynn D, Creevey C. RNA-seq analysis of differential gene expression in liver from lactating dairy cows divergent in negative energy balance. BMC Genomics. 2012;13(1):193.
11. McCarthy SD, Waters SM, Kenny DA, Diskin MG, Fitzpatrick R, Patton J, Wathes DC, Morris DG. Negative energy balance and hepatic gene expression patterns in high-yielding dairy cows during the early postpartum period: a global approach. Physiol Genomics. 2010;42(3):188–99.
12. Berry D, Crowley J. Residual intake and body weight gain: a new measure of efficiency in growing cattle. J Anim Sci. 2012;90(1):109–15.
13. Lin Z, Macleod I, Pryce J. Short communication: estimation of genetic parameters for residual feed intake and feeding behavior traits in dairy heifers. J Dairy Sci. 2013;96(4):2654–6.
14. Tempelman R, Spurlock D, Coffey M, Veerkamp R, Armentano L, Weigel K, De Haas Y, Staples C, Connor E, Lu Y. Heterogeneity in genetic and nongenetic variation and energy sink relationships for residual feed intake across research stations and countries. J Dairy Sci. 2015;98(3):2013–26.
15. Connor E, Hutchison J, Norman H, Olson K, Van Tassell C, Leith J, Baldwin R. Use of residual feed intake in Holsteins during early lactation shows potential to improve feed efficiency through genetic selection. J Anim Sci. 2013;91(8):3978–88.
16. Koch RM, Swiger LA, Chambers D, Gregory K. Efficiency of feed use in beef cattle. J Anim Sci. 1963;22(2):486–94.
17. Kennedy B, Van der Werf J, Meuwissen T. Genetic and statistical properties of residual feed intake. J Anim Sci. 1993;71(12):3239–50.
18. Pryce J, Arias J, Bowman P, Davis S, Macdonald K, Waghorn G, Wales W, Williams Y, Spelman R, Hayes B. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. J Dairy Sci. 2012;95(4):2108–19.
19. Connor E, Hutchison J, Olson K, Norman H. Triennial lactation symposium: Opportunities for improving milk production efficiency in dairy cattle. J Anim Sci. 2012;90(5):1687–94.
20. Do DN, Ostersen T, Strathe AB, Mark T, Jensen J, Kadarmideen HN. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. BMC Genet. 2014;15(1):1.

21. Do DN, Strathe AB, Ostersen T, Pant SD, Kadarmideen HN. Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. Front Genet. 2014;5:307.
22. Berry D, Crowley J. Cell biology symposium: genetics of feed efficiency in dairy and beef cattle. J Anim Sci. 2013;91(4):1594–613.
23. Partridge CG, Fawcett GL, Wang B, Semenkovich CF, Cheverud JM. The effect of dietary fat intake on hepatic gene expression in LG/J AND SM/J mice. BMC Genomics. 2014;15(1):1.
24. Kadarmideen HN. Genomics to systems biology in animal and veterinary sciences: progress, lessons and opportunities. Livest Sci. 2014;166:232–48.
25. Suravajhala P, Kogelman LJ, Kadarmideen HN. Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare. Genet Sel Evol. 2016;48(1):1.
26. Zhang ZH, Jhaveri DJ, Marshall VM, Bauer DC, Edson J, Narayanan RK, Robinson GJ, Lundberg AE, Bartlett PF, Wray NR. A comparative study of techniques for differential expression analysis on RNA-Seq data. PLoS One. 2014;9(8):e103207.
27. Drackley JK, Overton TR, Douglas GN. Adaptations of glucose and long-chain fatty acid metabolism in liver of dairy cows during the periparturient period. J Dairy Sci. 2001;84:E100–12.
28. Alexandre PA, Kogelman LJ, Santana MH, Passarelli D, Pulz LH, Fantinato-Neto P, Silva PL, Leme PR, Strefezzi RF, Coutinho LL. Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. BMC Genomics. 2015;16(1):1.
29. Yi G, Yuan J, Bi H, Yan W, Yang N, Qu L. In-Depth Duodenal Transcriptome Survey in Chickens with Divergent Feed Efficiency Using RNA-Seq. PLoS One. 2015;10(9):e0136765.
30. Kantartzis K, Machicao F, Machann J, Schick F, Fritsche A, Häring H-U, Stefan N. The DGAT2 gene is a candidate for the dissociation between fatty liver and insulin resistance in humans. Clin Sci. 2009;116(6):531–7.
31. Paradis F, Yue S, Grant J, Stothard P, Basarab J, Fitzsimmons C. Transcriptomic analysis by RNA sequencing reveals that hepatic interferon-induced genes may be associated with feed efficiency in beef heifers. J Anim Sci. 2015;93(7):3331–41.
32. Ozuna AC, Rowland RR, Nietfeld JC, Kerrigan MA, Dekkers J, Wyatt CR. Preliminary findings of a previously unrecognized porcine primary immunodeficiency disorder. Vet Pathol Online. 2012;2012:0300985812457790.
33. Kogelman LJ, Cirera S, Zhernakova DV, Fredholm M, Franke L, Kadarmideen HN. Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. BMC Med Genomics. 2014;7(1):1.
34. Do DN, Strathe AB, Ostersen T, Jensen J, Mark T, Kadarmideen HN. Genome-wide association study reveals genetic architecture of eating behavior in pigs and its implications for humans obesity by comparative mapping. PLoS One. 2013;8(8):e71509.
35. de Almeida Santana MH, Junior GAO, Cesar ASM, e Silva SL, Freua MC, da Costa Gomes R, Leme PR, Fukumasu H, Carvalho ME, Ventura RV. Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. J Appl Genet. 2016;2016:1–10.
36. Santana M, Ventura R, Utsunomiya Y, Neves H, Alexandre P, Oliveira Junior G, Gomes R, Bonin M, Coutinho L, Garcia J. A genomewide association mapping study using ultrasound-scanned information identifies potential genomic regions and candidate genes affecting carcass traits in Nellore cattle. J Anim Breed Genet. 2015;132(6):420–7.
37. Grasfeder LL, Gaillard S, Hammes SR, Ilkayeva O, Newgard CB, Hochberg RB, Dwyer MA, Chang C-y, McDonnell DP. Fasting-induced hepatic production of DHEA is regulated by PGC-1α, ERRα, and HNF4α. Mol Endocrinol. 2009;23(8):1171–82.
38. Chen Y, Gondro C, Quinn K, Herd R, Parnell P, Vanselow B. Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. Anim Genet. 2011;42(5):475–90.
39. Tizioto PC, Coutinho LL, Decker JE, Schnabel RD, Rosa KO, Oliveira PS, Souza MM, Mourão GB, Tullio RR, Chaves AS. Global liver gene expression differences in Nelore steers with divergent residual feed intake phenotypes. BMC Genomics. 2015;16(1):1.
40. da Costa ASH, Pires VMR, Fontes CMGA, Prates JAM. Expression of genes controlling fat deposition in two genetically diverse beef cattle breeds fed high or low silage diets. BMC Vet Res. 2013;9(1):1.
41. Sumner-Thomson J, Vierck J, McNamara J. Differential expression of genes in adipose tissue of first-lactation dairy cattle. J Dairy Sci. 2011;94(1):361–9.
42. Grala TM, Kay JK, Phyn CV, Bionaz M, Walker CG, Rius AG, Snell RG, Roche JR. Reducing milking frequency during nutrient restriction has no effect on the hepatic transcriptome of lactating dairy cattle. Physiol Genomics. 2013;45(23):1157–67.
43. Yu YH, Ginsberg H. The role of acyl-CoA: diacylglycerol acyltransferase (DGAT) in energy metabolism. Ann Med. 2004;36(4):252–61.
44. Chen HC, Farese RV. DGAT and triglyceride synthesis: a new target for obesity treatment? Trends Cardiovasc Med. 2000;10(5):188–92.
45. GeneCards. http://www.genecards.org/. Accessed 15 Mar 2017.
46. Li B, Fikse W, Lassen J, Lidauer M, Løvendahl P, Mäntysaari P, Berglund B. Genetic parameters for dry matter intake in primiparous Holstein, Nordic Red, and Jersey cows in the first half of lactation. J Dairy Sci. 2016;99(9):7232–9.
47. Andrews S: Fast QC. A quality control tool for high throughput sequence data, Reference Source. 2010.
48. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):10–2.
49. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
50. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292–4.
51. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics. 2014;2014:btu638.
52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):1.
53. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):1.
54. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34(3):267–73.
55. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005;102(43):15545–50.
56. Irizarry RA, Wang C, Zhou Y, Speed TP. Gene set enrichment analysis made simple. Stat Methods Med Res. 2009;18(6):565–75.
57. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2014;43(Database issue):D447–52.

Manuscript 2: Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency

---

Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency

Salleh, M.S.[1,2], Mazzoni, G. [1,2], Løvendahl, P.[3], and Kadarmideen, H.N.[1,2*]

[1] Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, DK-1870, Frederiksberg C, Denmark

[2] Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark

[3] Department of Molecular Biology and Genetics - Center for Quantitative Genetics and Genomics, Aarhus University, AU Foulum, DK-8830, Tjele, Denmark

* Corresponding author

SMS: surayams@sund.ku.dk

GM: gianmaz@bioinformatics.dtu.dk

PL: peter.lovendahl@mbg.au.dk

HNK: hajak@dtu.dk

Abstract

Background: Selection for feed efficiency is crucial for overall profitability and sustainability in dairy cattle production. Key regulator genes and genetic markers derived from co-expression networks underlying feed efficiency could be included in the genomic selection of the best cows. The present study identified co-expression networks associated with high and low feed efficiency and their regulator genes in Danish Holstein and Jersey cows.

RNA-sequencing data from Holstein and Jersey cows with high and low residual feed intake (RFI) and treated with two diets (low and high concentrate) were used. Approximately 26

million and 25 million pair reads were mapped to bovine reference genome for Jersey and Holstein breed, respectively. Subsequently, the gene count expressions data were analysed using a Weighted Gene Co-expression Network Analysis (WGCNA) approach. Functional enrichment analysis from Ingenuity® Pathway Analysis (IPA®), ClueGO application and STRING of these modules was performed to identify relevant biological pathways and regulatory genes.

Results: WGCNA identified two groups of co-expressed genes (modules) significantly associated with RFI and one module significantly associated with diet. In Holstein cows, the salmon module with module trait relationship (MTR)= 0.7 and the top upstream regulators ATP7B were involved in cholesterol biosynthesis, steroid biosynthesis, lipid biosynthesis and fatty acid metabolism. The magenta module has been significantly associated (MTR= 0.51) with the treatment diet involved in the triglyceride homeostasis. In Jersey cows, the lightsteelblue1 (MTR=-0.57) module controlled by IFNG and IL10RA was involved in the positive regulation of interferon-gamma production, lymphocyte differentiation, natural killer cell-mediated cytotoxicity and primary immunodeficiency.

Conclusion: The present study provides new information on the biological functions in liver that are potentially involved in controlling feed efficiency. The hub genes and upstream regulators (ATP7b, IFNG and IL10RA) involved in these functions are potential candidate genes for the development of new biomarkers. However, the hub genes, upstream regulators and pathways involved in the co-expressed networks were different in both breeds. Hence, additional studies are required to investigate and confirm these findings prior to their use as candidate genes.

Keywords: RNA-seq, Feed efficiency, Residual feed intake, Co-expressed genes, Hub genes, Pathways, Holstein, Jersey, Dairy cattle

Background

Globally, food demand is increasing as a consequence of world population growth [1]. However, arable land to produce sufficient amounts of food is decreasing, and the carbon footprint is increasing [2]. Hence, solutions for efficient and environmentally friendly methods to produce food are urgently needed.

Feed efficiency (FE) in dairy cattle is the ability of a cow to convert the feed nutrient consumed into milk and milk by-products. Many approaches have been developed and adopted to select the most feed-efficient cows. Currently, residual feed intake (RFI) has been used to measure FE in dairy cows [3, 4]. Residual feed intake is the difference between the predicted and actual feed intake [5]. Regression models have been used to calculate the RFI value. Thus, animals with low RFI values are more efficient [6]. The genetic selection of animals with a low RFI will improve profitability [7], decrease greenhouse gasses emissions [8] and optimize the use of food resources. However, in the case of dairy cattle, the interpretation of RFI is not straightforward. Many other factors should be considered, as this selection might lead to a negative energy balance, cause health issues and affect the fertility of the cows [9, 10].

In Denmark, Holstein and Jersey are the most common dairy breeds used [11]. Comparatively, Holstein and Jersey cattle do not differ in terms of digestibility, energy efficiencies, and the ability to convert dietary protein to milk protein [12]. However, there are no gene expression

63 profiling studies of these breeds. Hence, to understand the complex biological mechanisms in
64 nutrient partitioning in dairy cattle, liver transcriptomics analysis may be useful to interpret and
65 understand the pathways and functional elements of the genomes involved[13].
66 Transcriptomics is a form of high throughput analysis to quantify gene expression in a specific
67 cell type or tissue [14]. Various studies have reported that mRNA levels of many genes are
68 heritable, which affects genetic analysis [15-17]. Many studies based on transcriptomics
69 (microarray and RNA-sequencing) have been conducted to study gene expression in feed
70 efficiency [18-20]. Studies on differential gene expression have been well established to identify
71 candidate genes for biomarker development [21]. There are limited studies related to gene
72 expression for RFI traits in dairy cattle, particularly for Jersey and Holstein breeds. However,
73 some studies have reported the gene expression associated with RFI in other breeds and
74 species. For example, Lkhagvadorj et al. (2010) [22] found that the common energy
75 consumption controlled by *PPARA*, *PPARG* and/or *CREB* is related to RFI in pigs. In beef cattle,
76 Alexandre et al. (2015)[19] reported the alteration of lipid metabolism and an increase in the
77 inflammatory response in animals with low feed efficiency. Paradis et al. (2015) [20] also
78 reported a greater response to hepatic inflammation in heifers with high feed efficiency. In
79 Nellore beef cattle, Tizioto et al., 2016 [23] identified the differentially expressed genes
80 involved in oxidative stress. Hence, transcriptomics analysis might provide additional
81 knowledge on the complex mechanisms that regulate nutrient intake.

82 Diet affects the energy metabolism and efficiency of dairy cows [24]. Some studies have
83 investigated the correlation between FE and diet, focusing on the gene expression profiles of
84 specific tissues. Dairy cows are typically fed high energy or high-concentrate feed to meet the

high-energy demand during the lactation period. It has previously been shown that high energy feeding does not affect the fatty acid concentration but does affect the expression of genes such as *ACACA*, *LPL* and *SCD* in the lipid metabolism [25]. Thus, it is also interesting to investigate the effects of different levels of energy in feed using co-expression network approaches.

Previously, we performed differential gene expression analysis on RNA from the livers of Holstein and Jersey cows. We identified several differentially expressed genes between high and low RFI [26]. The differentially expressed genes were related to primary immunodeficiency, steroid hormone biosynthesis, retinol metabolism, starch and sucrose metabolism, ether lipid metabolism, arachidonic metabolism and cytochrome P450 in drug metabolism. These biological processes and pathways are important mechanisms that are associated with feed efficiency.

Therefore, it is important to thoroughly investigate the mechanisms controlling feed efficiency. Systems biology is the most promising approach to obtain a better understanding of complex traits, such as feed efficiency. In systems biology, many computational methods are based on network approaches. Co-expression network analysis has been successfully used to analyse complex traits and diseases in humans and animals [27-30]. Weighted Gene Co-expression Network Analysis (WGCNA) can be used to identify clusters (modules) of highly correlated genes [31]. WGCNA has been used to identify candidate genes that are associated with the FE. Alexandra et al. (2015) identified differentially co-expressed genes that are involved in lipid

105 metabolism in RFI divergent Nellore cattle. Similarly, lipid metabolism-related processes were
106 identified in low-RFI pigs [22].

107 In the present study, the WGCNA method was applied to RNA-Seq data from the livers of
108 Holstein and Jersey cows to: i) identify groups of co-expressed genes and biological pathways
109 associated with RFI; ii) identify the hub genes and upstream regulators in these modules that
110 may be good candidate genes for feed efficiency-related traits; and iii) compare the
111 mechanisms and processes involved in RFI between Holstein and Jersey cattle. To our
112 knowledge, this study is the first to use weighted gene network approaches to examine the
113 overall complex transcriptional regulation of feed efficiency (RFI) using RNA-Seq data in Danish
114 Holstein and Jersey cows.

115 ## Materials and Methods

116 **Animal ethics statement**

117 The experimental design and animals that were being used in this experiment were permitted
118 by the Danish Animal Experimentation Inspectorate.

119 **Experimental data**

120 The experimental design and details of the experimental animals have been previously
121 described in [26].

122 In brief, the dataset used in this experiment consists of 38 RNA-Seq expression profiles of liver
123 bioposies from nine Holsteins and ten Jersey cows. In each breed group, cows were classified in
124 high and low feed efficient and RNA samples were collected before and after treatment diet

(low and high concentrate diet). The animals were assigned to the different diets after at least for 14-26 days adaptation period. All 38 RNA samples were paired-end sequenced using Illumina HiSeq 2500. The bioinformatics pipeline for RNA-Seq data processing is described in [26]. The expression quantification was performed using Ensembl Bovine annotation (release 82). The raw count data matrix used in this study is available in http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92398.

Weighted Gene Co-expression Network Analysis (WGCNA)

The Weighted Gene Co-expression Network Analysis (WGCNA) [31] R package was used to build co-expression networks and identify groups of highly co-expressed genes. Individual analyses were conducted on each breed group.

First, the low count genes and outliers were filtered by leaving only genes that had at least 1 count per million in 90% of the group. The remaining 11,153 genes in Holstein and 11,238 genes in Jersey were used for the analysis. The gene expression counts were normalized using the default procedure from the DESeq2 package version 1.12.0 [32] by correcting for the parity number to reduce potential effects from the parity number factor. The normalized data were subsequently log transformed as suggested in the WGCNA manual [33]. The final dataset was used in WGCNA to build an unsigned network. Pairwise Pearson's correlations among all genes were calculated to create an adjacency matrix. A soft threshold power was set at $\beta = 12$ for Holstein and $\beta = 10$ for Jersey, correspondent to a scale-free topology index (R^2) [34] of 0.9 for Holstein and 0.8 for Jersey. The adjacency matrix was used to calculate the Topological Overlap

Measure (TOM). Modules of co-expressed genes were identified by using the dynamic tree cut algorithm [35]. Modules were arbitrarily labelled with different colours.

The module eigengenes were computed for each module using the first principal component to capture the variation in gene expression within each module. The correlation between module eigengene and RFI or treatment diet was evaluated to select modules that were associated with the respective traits (p-value <0.05). Gene significance (GS) was computed for each gene as the correlation between gene expression counts and FE. In addition, hub genes were identified, selecting genes with high module membership (MM> 0.8) in the modules of interest.

Functional enrichment analysis

The modules that are significantly associated with RFI and treatment diet traits were selected.

Functional enrichment analysis was performed in the selected modules to identify and interpret complex biological functions based on gene ontology terms for the biological processes, molecular functions and cellular components and based on the KEGG pathways annotation.

All the genes included in each module were used in the functional enrichment analysis with the Cytoscape 3.4.0 plug-in software, ClueGO v2.2.6 [36]. The significance value was set as p-value <0.05 and the Benjamini–Hochberg (BH) correction was used as the multiple test correction. The reference set used for this analysis included a total of 9064 genes. The list of genes in the module of interest was also analysed using the STRING v.10.0 [37] database and the *Bos taurus* annotation.

Ingenuity® Pathway Analysis (IPA®) was used to detect upstream regulators, diseases and functions in the selected modules.

A summary of the pipeline of the experimental workflow, bioinformatics and statistical analysis is presented in **Figure 1**.

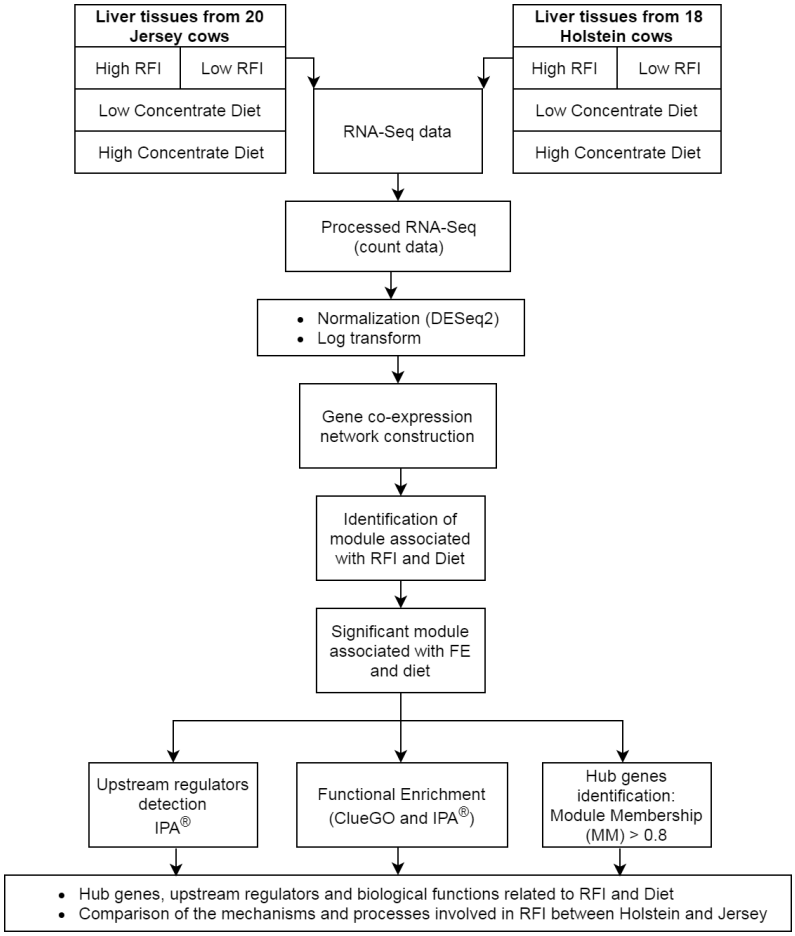
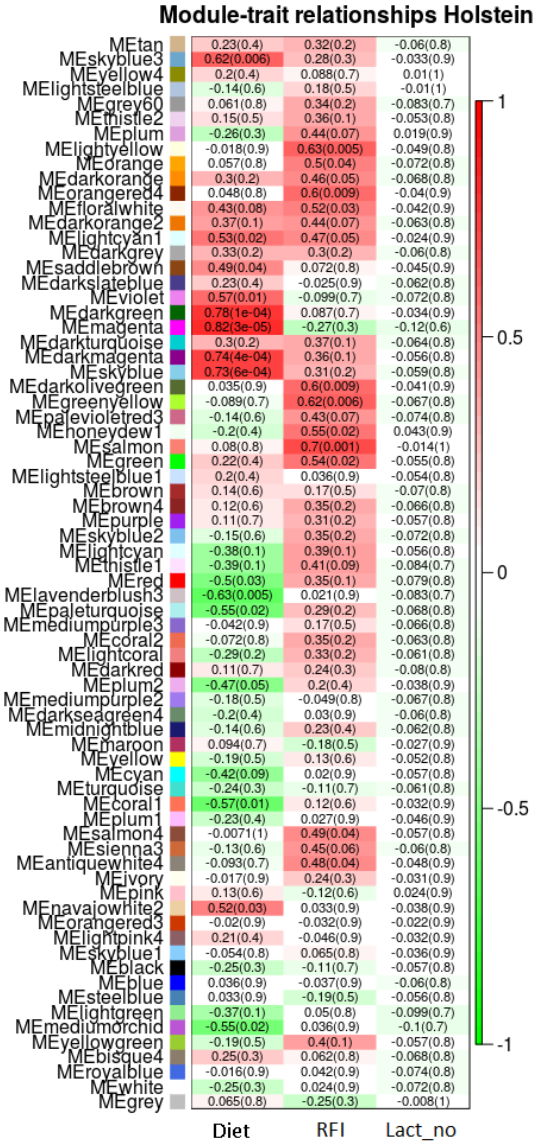


Figure 1 Experimental design and co-expressed gene network analysis pipeline

Results

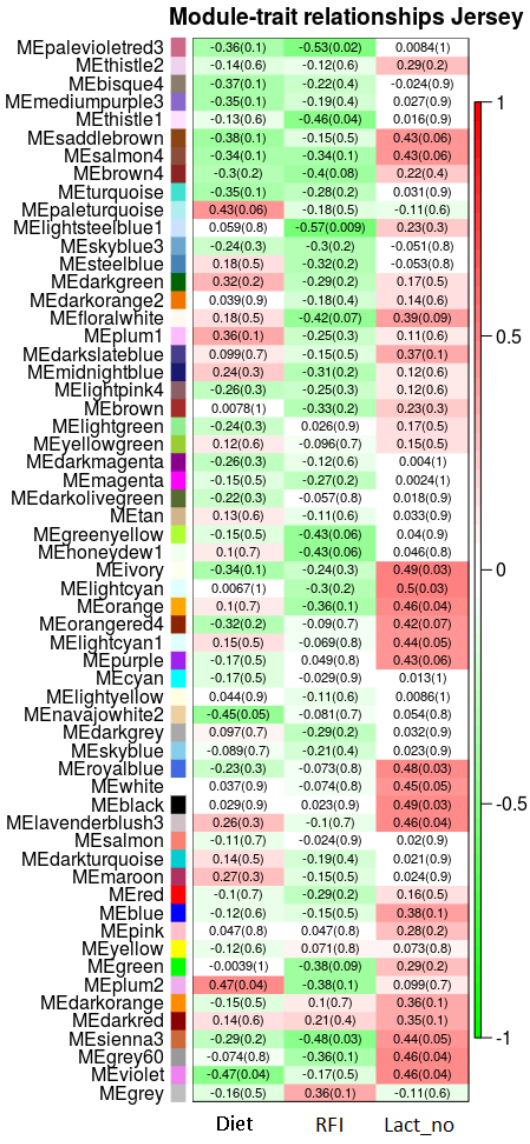
In the present study, WGCNA was used to identify RFI and diet-associated co-expression modules and their key functions. In total, 72 modules (**Figure 2**) for Holstein cows and 59 modules (**Figure 5**) for Jersey cows were identified. A total of 13 modules and three modules were significantly correlated with RFI for Holstein and Jersey cows, respectively. Additionally, 14 modules for Holstein and three modules for Jersey were significantly associated with treatment diet.

We assigned all the significant modules into the ClueGO application analysis to investigate the gene ontology (GO) and KEGG pathway-related functions with specific traits. The modules with the most significant module trait relationships (MTRs) and relevant biological information were selected as the modules of interest in the present study. The modules lightsteelblue1 and violet in Jersey cows and the modules salmon and magenta in Holstein cows were selected for RFI and treatment diet, respectively.



184

185 **Figure 2** Module trait relationship (p-value) for detected modules (y-axis) in relation with
186 traits (x-axis) for Holstein cows. The module trait relationship were colored based on the
187 correlation between the module and traits (red=strong positive correlation; green=strong
188 negative correlation). X-axis legend: Diet= Treatment diet; RFI= Residual feed intake; Lact_no=
189 Lactation number



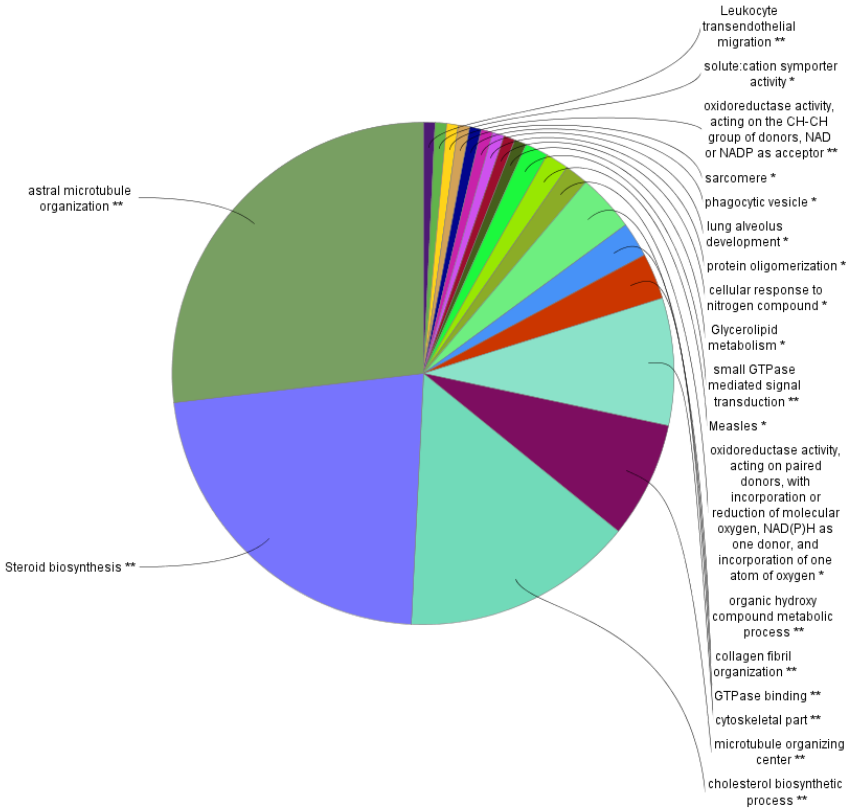
190

191 **Figure 5** Module trait relationship (p-value) for detected modules (y-axis) in relation with
192 traits (x-axis) for Jersey cows. The module trait relationship were colored based on the
193 correlation between the module and traits (red=strong positive correlation; green=strong
194 negative correlation). X-axis legend: Diet= Treatment diet; RFI= Residual feed intake; Lact_no=
195 Lactation number

Modules related to RFI and treatment diet in Holstein cows

In Holstein cows, 72 modules were significantly (p-value< 0.05) related to the salmon module (181 genes with MTR RFI= 0.7) as the most significant module associated with RFI. For the diet trait, we identified the magenta module as the most significant module. The magenta module comprised 212 genes that contribute to the MTR Diet= 0.82.

In the top module (salmon), steroid biosynthesis was identified as the most enriched KEGG pathway (**Figure 3**). This finding was also confirmed after analysing the genes in this module using www.string-db.org, and almost the same pathways and same patterns appeared in the output. Interestingly, most of the enriched pathways of co-expressed genes in Holstein cows were involved in steroid, lipid and cholesterol biosynthesis and metabolism (**Figure 3**).



206

207 **Figure 3** Pie chart presenting an overview of the significant GO terms and KEGG pathways
208 in the salmon module in Holstein cows

209 **Supplementary file, table 1** shows a summary of the functional groups with the number of

210 genes involved in the GO terms and pathways. In total, 84 GO terms were significantly enriched

211 (p-value< 0.05) after multiple testing correction using BH. The GO-terms and KEGG pathways

212 presented here are also almost the same as the output from the STRING 10 analysis

213 (**supplementary file, tables 3, 4 and 5**).

214 The list of upstream regulators identified for the modules that are significantly associated with

215 RFI and diet are presented in **supplementary file, table 9.** In the salmon module, ATP7B was

216 predicted as activated, while POR and cholesterol were predicted as inhibited. In

217 **supplementary file, table 11 and 12** shows the diseases and functions involved in salmon and

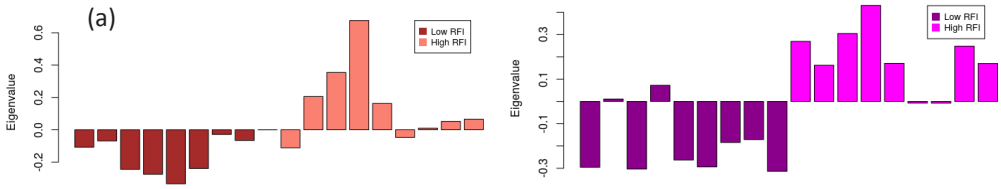
218 magenta modules.

219 A module eigengene diagram for both the salmon and magenta modules shows overexpression

220 in high RFI individuals (**Figure 4 a and b**).

221 The list of genes with high (MM> 0.8) in the salmon (b) e is presented in **table 1**.

222



223

224 **Figure 4** **(a)** Module eigengene (y-axis) across samples (x-axis) from the salmon module
225 (associated to RFI) **(b)** Module eigengene (y-axis) across samples (x-axis) from the magenta
226 module (associated to treatment diet)

227 **Table 1** List of the top hub genes generated from (MM> 0.8) in the salmon module in
228 Holstein cows

Ensembl gene ID	Gene name	Module membership	Gene significance
ENSBTAG00000000197	*TRMT10A*	0.801	0.576
ENSBTAG00000001774	*SPRY2*	-0.814	-0.520
ENSBTAG00000001950	*RDH11*	0.852	0.441
ENSBTAG00000002412	*CYB5B*	0.907	0.633
ENSBTAG00000002435	*PTPRE*	0.852	0.767
ENSBTAG00000002714	*GNAI1*	0.901	0.557
ENSBTAG00000002827	*ACAT2*	0.946	0.691
ENSBTAG00000002966	*DNAJC13*	0.813	0.710
ENSBTAG00000003068	*MSMO1*	0.852	0.579
ENSBTAG00000003305	*NCF1*	0.802	0.642
ENSBTAG00000003696	*CCDC64*	0.837	0.679
ENSBTAG00000003718	*HACL1*	0.854	0.705

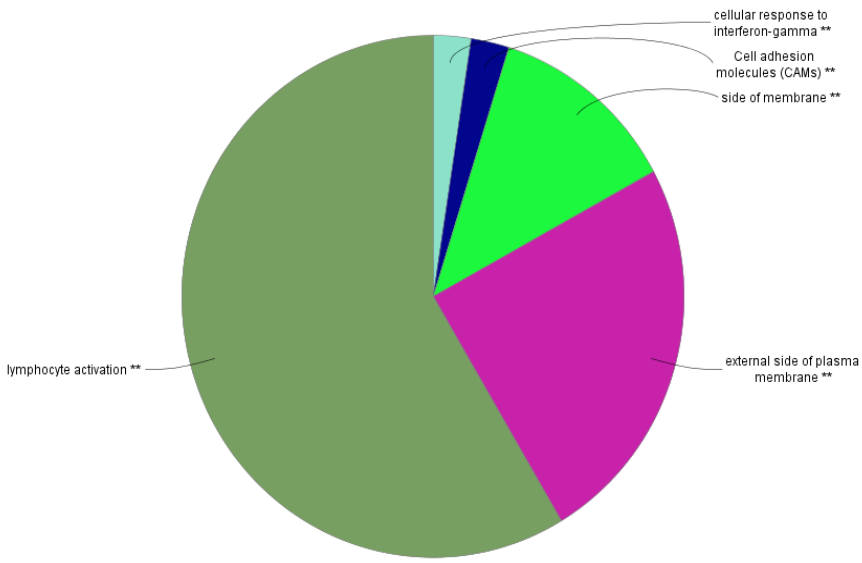
ENSBTAG00000003948		0.919	0.559
ENSBTAG00000004075	*IDI1*	0.870	0.607
ENSBTAG00000004688	*DHCR24*	0.859	0.555
ENSBTAG00000005183	*MVK*	0.906	0.497
ENSBTAG00000005498	*SQLE*	0.816	0.442
ENSBTAG00000005650	*SKAP2*	0.826	0.589
ENSBTAG00000005976	*HSD17B7*	0.809	0.550
ENSBTAG00000006999	*RYR1*	0.929	0.763
ENSBTAG00000007014	*CEP63*	0.823	0.623
ENSBTAG00000007079	*LCP1*	0.806	0.583
ENSBTAG00000007840	*HMGCR*	0.888	0.522
ENSBTAG00000007844	*CETN2*	0.836	0.335
ENSBTAG00000008160	*MBOAT2*	0.865	0.534
ENSBTAG00000008329	*CYTIP*	0.823	0.477
ENSBTAG00000010347	*EZR*	0.850	0.506
ENSBTAG00000011146	*RAB8B*	0.884	0.473
ENSBTAG00000011839	*HMGCS1*	0.871	0.507
ENSBTAG00000012059	*MVD*	0.831	0.364
ENSBTAG00000012170	*UBL3*	0.813	0.729
ENSBTAG00000012432	*FDFT1*	0.821	0.529
ENSBTAG00000012695	*LCK*	0.837	0.534
ENSBTAG00000013284		0.886	0.736
ENSBTAG00000013303	*ACSS2*	0.866	0.571
ENSBTAG00000013749	*RHOQ*	0.868	0.525
ENSBTAG00000014517	*KLB*	0.857	0.640
ENSBTAG00000015327	*SPTAN1*	0.899	0.637
ENSBTAG00000015980	*FASN*	0.859	0.490
ENSBTAG00000016445	*YME1L1*	0.807	0.717
ENSBTAG00000016465	*DHCR7*	0.903	0.521
ENSBTAG00000016709	*NT5C3A*	0.824	0.615
ENSBTAG00000016721	*ZNF791*	0.824	0.559
ENSBTAG00000016740	*ACLY*	0.918	0.520
ENSBTAG00000018936	*LSS*	0.839	0.580
ENSBTAG00000018959	*RAB11A*	0.828	0.670
ENSBTAG00000020984	*RAPGEF4*	0.856	0.775
ENSBTAG00000021842		0.804	0.492
ENSBTAG00000030951		0.844	0.508
ENSBTAG00000036260	*LPXN*	0.801	0.391
ENSBTAG00000037413	*TMEM164*	0.810	0.468
ENSBTAG00000047970		0.835	0.558

229

Modules related to RFI and treatment diet in Jersey cows

Among the 14 modules significantly (p-value< 0.05) related to RFI in the Jersey group, the lightsteelblue1 module (57 genes) with a module trait relationship (MTR RFI= -0.57) is the most significant (p-value< 0.05) module associated with RFI. In total, 44 GO terms were significantly enriched (p-value< 0.05) after multiple test correction using BH. For the diet trait, among the three significantly correlated modules, the violet module was the most significant (MTR Diet= -0.47). However, this module has limited output from a functional enrichment analysis or no interesting biological information related to diet. Hence, the modules related to diet for the Jersey breed were not further discussed.

Figure 6 and **supplementary file, table 2 shows** the top summarized GO terms involved in the lightsteelblue1 module that is related to immune system functions. The first and the second GO terms, which are associated with the regulation of lymphocyte activation and positive regulation of leukocyte activation, involved almost the same genes as those that are involved in immune system functions. In detail, primary immunodeficiency has been identified (p-value< 0.05) as a significant KEGG pathway that involves four genes together with the positive regulation of leukocyte activated GO terms.



246

247 **Figure 6** Pie chart visualization of GO terms and KEGG pathways in the lightsteelblue1
248 module in Jersey cows.

249 We identified IFNG (Interferon Gamma) as inhibited and IL10RA (Interleukin 10 Receptor
250 Subunit Alpha), NKX2-3 (NK2 Homeobox 3) and dexamethasone were predicted as activated
251 upstream regulators (**Supplementary file, table 10).** In **supplementary file ,table 13 and 14**
252 shows the diseases and functions involved in lightsteelblue1 and violet modules.

253 Interestingly, all of these upstream regulators have functions related to the immune system. In
254 addition, GO-terms and KEGG pathways from the STRING 10 analysis (**supplementary file,**
255 **tables 6, 7 and 8**) also give almost the same output.

256 The module eigengene for the lightsteelblue1 module shows an under-expression profile in
257 high RFI individuals (**Figure 7**).

The list of genes with high (MM> 0.8) in the lightsteelblue1 module is presented in **table 2**.

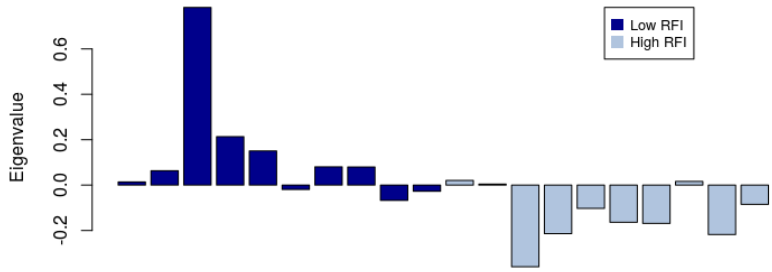


Figure 7 Module eigengene (y-axis) across samples (x-axis) from the lightsteelblue1 module (associated to RFI)

Table 2 List of the top hub genes generated from (MM> 0.8) in the lightsteelblue1 module in Jersey cows

Ensembl gene ID	Gene name	Module membership	Gene significance
ENSBTAG00000000431	*TRDC*	0.858	-0.411
ENSBTAG00000000432	*TRAC*	0.860	-0.526
ENSBTAG00000000715		0.889	-0.487
ENSBTAG00000001198		0.810	-0.555
ENSBTAG00000002669	*RASSF4*	0.802	-0.722
ENSBTAG00000003037		0.829	-0.485
ENSBTAG00000004894		0.907	-0.497
ENSBTAG00000004917	*KLRK1*	0.826	-0.437
ENSBTAG00000005628		0.818	-0.490
ENSBTAG00000005892	*ZAP70*	0.864	-0.609
ENSBTAG00000006452	*CD3D*	0.900	-0.494
ENSBTAG00000006552	*LAMP3*	0.827	-0.501
ENSBTAG00000007191	*CCL5*	0.909	-0.480
ENSBTAG00000008401	*PFKFB3*	0.808	-0.547
ENSBTAG00000009381	*LCP2*	0.857	-0.654
ENSBTAG00000012695	*LCK*	0.852	-0.510
ENSBTAG00000013730	*CD5*	0.857	-0.403
ENSBTAG00000014725	*CD27*	0.822	-0.474
ENSBTAG00000015708	*CXCR6*	0.879	-0.469
ENSBTAG00000015710	*CD3E*	0.875	-0.537
ENSBTAG00000017256	*CD2*	0.914	-0.474
ENSBTAG00000019403	*MALSU1*	0.800	-0.536

ENSBTAG00000020904	*JAK3*	0.857	-0.439
ENSBTAG00000027246	*UBD*	0.888	-0.621
ENSBTAG00000030426		0.889	-0.379
ENSBTAG00000037510		0.853	-0.433
ENSBTAG00000038639	*CXCL9*	0.906	-0.425
ENSBTAG00000039588		0.815	-0.535
ENSBTAG00000047988		0.842	-0.365

Discussion

WGCNA identified groups of co-expressed genes that are expected to perform the same biological functions and affect RFI. From the MTR, we tested the modules that were significantly correlated to the focus traits (RFI and diet). However, only the most significant module had any interesting biological meaning associated with the traits (one module in each breed). Hence, only the most biologically meaningful modules were further analysed and discussed.

For Holstein cows, we identified pathways and upstream regulators related to steroid biosynthesis, lipid metabolism, cholesterol metabolism and production in salmon module. In particular, we identified the activation of cholesterol and lipid synthesis in high RFI cows. There was a tendency for these three mechanisms to be activated in the datasets, which is consistent with the idea that high synthesis of fat is correlated with the loss of energy used in milk production in dairy cows, resulting in less feed efficient animals [38]. This finding is also consistent with previous studies that associated high fat deposition with high RFI animals [6, 39]. The magenta module was significantly associated with diet and involved the energy consumption and regulation of glucose.

For Jersey cows, the lightsteelblue1 module was enriched for immune system-related functions. Interestingly, the upstream regulators for the genes in the lightsteelblue1 module (IFNG and IL10RA) were also related to the immune system. In particular, the immune system in high RFI group was activated. Thus, the activation of the immune system leads to low feed efficiency, which is consistent with previous studies [19, 40].

These findings are supported by evidence from the co-expression network analysis of both breeds.

Co-expressed networks in Holstein cows

The functional enrichment analysis determined that the module identified in Holstein cows was involved in cholesterol biosynthesis, steroid biosynthesis , lipid biosynthesis and fatty acid metabolism.

From the most significant pathways, cholesterol biosynthesis has previously been discussed, as its related genes are important in the RFI. The cholesterol biosynthetic pathway is responsible for the variability of cholesterol levels in cells [41]. This module was also enriched for lipid biosynthesis. Interestingly, the levels of cholesterol and lipids have previously been positively associated with RFI in beef cattle [42].

Many genes in this modules have previously been associated with feed efficiency, [40]. For example, Acetyl-CoA carboxylase alpha (*ACACA*), Acetyl-CoA Acetyltransferase 2 (*ACAT2)*, and fatty acid synthase (*FASN)* genes in the modules are key genes in cholesterol biosynthesis, organic hydroxy compound metabolism, collagen fibril organization, steroid biosynthesis, astral microtubule organization, protein oligomerization and oxidoreductase activity, acting on the

CH-CH group of donors and NAD or NADP as an acceptor. *ACACA* and *FASN* were found to be differentially expressed and co-expressed in other feed efficiency-related studies [22, 40, 43]. The main function of *FASN* is to catalyse the synthesis of palmitate from acetyl-CoA and malonyl-CoA, in the presence of NADPH, into long-chain saturated fatty acids. Hence, these genes have a tendency to affect the feed efficiency in Holstein cows. In addition, many studies have discussed the involvement of several genes included in the modules that we identified in the present study (*CYP7A1, ACACA, FASN*) [40, 44]. The presence of *ACAT2* is also interesting because the product of this gene is involved in lipid metabolism [45].

Other feed efficiency studies, for example, in pigs, have previously observed that lipogenesis and steroidogenesis in liver tissue are closely related to feed efficiency [22, 46], confirming previous observations in the differential expression analysis of this dataset [26].

In Holstein cows, we identified ATP7B as a top upstream regulator for the salmon module. This protein uses energy in the molecule adenosine triphosphate (ATP), which is responsible for the transport of metals into and out of cells using the energy stored in the molecule adenosine triphosphate (ATP). ATP7B appears to be activated in high RFI (low FE). Hoogeveen et al. (1995) [47] stated that the deficiency of copper in rats would increase the utilization of fat in rats. Hence, this finding suggests a relationship when ATP7B is activated, which potentially reflects the deposition of fats. Consistent with the present study, the high RFI cow shows the activation of ATP7B. This upstream regulator shows a relationship with regulating the fat consumption. Although it is not straightforward, the presence of the gene reflects the consumption of fat and indirectly affects the fat composition [48].

In the present study, cholesterol synthesis was activated in the IPA upstream regulator analysis. Furthermore, the activation of lipid metabolism in the disease function analysis supports the evidence from the GO term and pathway analyses. As lipid and cholesterol metabolism, and fat synthesis in particular, are activated in the high RFI group, we can assume that the high RFI group is inefficient in converting fat to energy. Hence, animals with high RFI (low FE) have high levels of cholesterol and fat in the body [49]. This finding is also consistent with Arthur et al. (2001) [50], who reported the positive relationship between RFI and average back fat in beef carcasses.

Interestingly, when fed a high or low concentrate diet, triglyceride homeostasis was the top GO biological process, which might be the result of the high energy or low energy diet. A previous study reported that controlled diet (with fructose and glucose) significantly affects the triglyceride levels [51].

Generally, based on the results obtained from the functional enrichment analysis for the Holstein breed, the most important GO terms, KEGG pathways and upstream regulators involved were related to steroid biosynthesis, cholesterol biosynthesis, lipid biosynthesis and triglyceride homeostasis. These findings show that the feed efficiency in Holstein cows is strictly associated with the regulation of energy via lipid and cholesterol metabolism.

Co-expressed networks in Jersey cows

The most significant pathways in Jersey cows were positive regulation of interferon-gamma production, lymphocyte differentiation, side of membrane, natural killer cell-mediated cytotoxicity and primary immunodeficiency. Interestingly, these most summarized pathways

were related to the immune system. From the IPA upstream regulator and diseases function analysis, the immune system related functions were activated in the high RFI group.

Several studies also suggested that the involvement of the immune system would affect the feed efficiency [52, 53]. For example, [19, 27] discussed important findings but in different species and breeds. Kristina et al. (2016) [40] discovered an increase in the inflammatory response of the progeny of low RFI sires, which is consistent with the results of the present study. The type of diet might also affect the immune response. For example, Ametaj et al. (2009) [54] reported that the feeding of high concentrate feeds affects several inflammatory responses in feedlot steers. However, in the present study, no significant effect from the different type of concentrate diet in Jersey cows was observed. This finding might reflect the different populations and different breeds, as dairy cattle convert their nutrients into different products with respect to beef cattle[55]. Although, many other studies relate their findings with the importance of the immune system in RFI and feed efficiency, few studies have been conducted in dairy cattle [19, 20, 40].

Furthermore, these significant GO terms and pathways were also supported by the findings from upstream regulator analysis through IPA®. The top upstream regulator in Jersey cows is Interferon Gamma (IFNG), which has an interesting relationship to interactions among nutrition, metabolism, and the immune system [56]. This gene encodes a soluble cytokine that is a member of the type II interferon class. IFNG was predicted to be inhibited in high RFI Jersey cows. This protein is secreted from cells of both the innate and adaptive immune systems. IFNG is important in the system because it directly inhibits viral replication. The down-regulation of

this cytokine in the high RFI group in Jersey cows might affect the feed efficiency. Thus, IFNG plays an important role in regulating immune systems in animals. Another interesting upstream regulator in Jersey cows is IL10RA (Interleukin 10 Receptor Subunit Alpha), which was predicted to be activated. *IL10RA* is a receptor with anti-inflammatory properties [57]. The activation of this gene might result in inhibition of the synthesis of pro-inflammatory cytokines. Reynolds et al., 2017 [58] reported that IL10RA was differentially expressed in rumen papillae of divergent average daily gain steers and these authors showed a negative association between the inflammatory response and feed efficiency. Thus, the activation of IL10RA in the high RFI group would reflect the inflammatory response in Jersey cows.

We further speculate that, based on the results obtained in the Jersey breed, the most important GO terms, KEGG pathways and upstream regulators were related to the immune system. Jersey cows have many co-expressed genes that relate to the immune system to regulate feed utilization. It is likely that in Jersey cows, immunity plays a key role in substituting feed nutrient into milk and milk components. The immune response plays an important role in energy balance during milk production in dairy cows.

Comparison of RFI associated modules between Holstein and Jersey cows

In the datasets analysed in the present study, the most significant module associated with RFI differed between the Holstein and Jersey breeds. Furthermore, these modules were enriched for different sets of biological processes. This evidence suggests that the Holstein cow system is more reactive towards steroid biosynthesis, while Jersey cows have more reactions in their immune systems. Several studies have reported the importance of the lipid and cholesterol

386 metabolism and immune system related functions in feed efficiency traits in farm animals, likely
387 reflecting the complex role of the liver in regulating the nutrient uptake[59].

388 The hub genes of the modules identified in the present study represent potential candidate
389 genes for RFI. These findings might provide additional information and new insights into the
390 biological processes that are associated with RFI in these two main dairy breeds. Thus, we
391 speculated that in this study population, the liver transcriptomics profiles of the two main dairy
392 breeds are involved in two different biological processes. However, a comparative feed
393 efficiency study reported similar results in terms of digestibility and ratios of milk to body
394 weight and feed intake between Holstein and Jersey cows [12]. The sample sizes for gene
395 expression studies to achieve the same statistical power as genetic (GWAS type) studies are
396 typically substantially lower, often five to ten samples. However, the sample size of the present
397 study did not enable confirmation of whether the identified biological processes are breed
398 specific. To confirm this notion, the set of genes should be validated in other cows using qPCR
399 to confirm whether the expression patterns conform to different RFI-diet groups, which is out
400 of the scope of the present study.

401 **Conclusion**

402 In conclusion, the co-expression network analysis revealed important genes and pathways in
403 the liver that are involved in feed efficiency (RFI). In Holstein cows, the overall results showed
404 that genes and upstream regulators such as ATP7b in RFI-associated modules that were co-
405 expressed were primarily related to steroid and lipid biosynthesis. The results show that high
406 RFI Holstein cows have a high lipid and cholesterol metabolism. The co-expressed genes

associated with treatment diet were involved in triglyceride homeostasis. We observed different patterns of co-expressed genes involved in Jersey cows for which most of the co-expressed genes associated with RFI were related to the immune system in the most significant module. The upstream regulators IFNG and ILR10 that were predicted to be inhibited and activated, respectively, were closely associated with the immune system in Jersey cows. A high RFI Jersey cow tends to have a higher response to inflammation. The information of the functional enrichment from the analysis of co-expressed genes provides a better understanding of the mechanisms controlling RFI in Holstein and Jersey cows. Thus, the present study paves the way for the development of biomarkers for feed efficiency in dairy cattle.

List of abbreviations

RFI: Residual Feed Intake
RNA-seq: RNA-sequencing
WGCNA: Weighted Gene CO-expression Network Analysis
IPA®: Ingenuity® Pathway Analysis
MTR: Module Trait Relationship
FE: Feed Efficiency
mRNA: Messenger RNA
LC: Low Concentrate
HC: High Concentrate
DM: Dry Matter
cDNA: Complementary DNA
TOM: Topological Overlap Measure
GS: Gene Significance
MM: Module Membership
KEGG: Kyoto Encyclopedia of Genes and Genomes
BH: Benjamini-Hochberg
GO: Gene Ontology

Declarations

Ethics approval and consent to participate

The experimental design and animals that were being used in this experiment were permitted by the Danish Animal Experimentation Inspectorate.

Consent to publish

Not applicable.

Availability of data and materials

The data discussed in this publication were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE92398 at:

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92398

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the “Feed Utilization in Nordic Cattle” from Danish Milk Levy Foundation, Skejby Denmark. SMS received a stipend for PhD studies from the Ministry of Higher Education of Malaysia and Universiti Putra Malaysia and a tuition fee waiver from the University of Copenhagen.

Authors' contributions

HNK was the overall project leader who conceived and conducted this selective transcriptomics profiling study on cows with high/low feed efficiency, and supervised SMS in the laboratory

work and bioinformatics/systems biology analyses. PL provided feed efficiency measurements for the cattle in the experiment. SMS processed liver tissue for the RNA isolation and quality control of the RNA samples prior to RNA Sequencing. SMS analysed the data and drafted the original manuscript with assistance from GM. GM substantially contributed to the bioinformatics analyses. All authors wrote, read, and approved the final version of the manuscript.

Acknowledgements

We gratefully acknowledge collaborator scientists, specifically Peter Lund, Johanna Höglund and Dana Olijhoek, for contributions to the present study. We also thank the field staff and laboratory technicians who assisted with the experiments in the present study.

References

1. Carolan M: **Food Security and Policy**. 2017.
2. Food, Nations AOotU: **Food Wastage Footprint: Impacts on Natural Resources: Summary Report**: FAO; 2013.
3. Connor EE, Hutchison JL, Norman HD: **Estimating feed efficiency of lactating dairy cattle using residual feed intake**. *Feed Efficiency in the Beef Industry* 2012:159-173.
4. Varga G, Dechow C: **Can We Use Residual Feed Intake to Enhance Dairy Production Efficiency?** In: *Proceedings of the 22nd Tri-State Dairy Nutrition Conference, Fort Wayne, Indiana, USA, 23-24 April 2013: 2013*: Michigan State University; 2013: 131-140.
5. Koch RM, Swiger LA, Chambers D, Gregory KE: **Efficiency of feed use in beef cattle**. *Journal of animal science* 1963, **22**(2):486-494.
6. Herd R, Arthur P: **Physiological basis for residual feed intake**. *Journal of animal science* 2009, **87**(14_suppl):E64-E71.
7. Williams Y, Pryce J, Grainger C, Wales W, Linden N, Porker M, Hayes B: **Variation in residual feed intake in Holstein-Friesian dairy heifers in southern Australia**. *Journal of Dairy Science* 2011, **94**(9):4715-4725.
8. Arthur JP, Herd R: **Residual feed intake in beef cattle**. *Revista Brasileira de Zootecnia* 2008, **37**(SPE):269-279.

9. Veerkamp R, Koenen E, De Jong G: **Genetic correlations among body condition score, yield, and fertility in first-parity cows estimated by random regression models**. *Journal of Dairy Science* 2001, **84**(10):2327-2335.
10. Llewellyn S, Fitzpatrick R, Kenny D, Murphy J, Scaramuzzi R, Wathes D: **Effect of negative energy balance on the insulin-like growth factor system in pre-recruitment ovarian follicles of post partum dairy cows**. *Reproduction* 2007, **133**(3):627-639.
11. **RYK- Ydelseskontrol, https://www.landbrugsinfo.dk/Kvaeg/RYK/Sider/Avereagemilkyieldisnowcloseto10000kgprcowandyears.aspx, Accessed date: 12-06-2017**.
12. Blake RW, Custodio AA, Howard WH: **Comparative Feed Efficiency of Holstein and Jersey Cows1**. *Journal of dairy science* 1986, **69**(5):1302-1308.
13. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics**. *Nature reviews genetics* 2009, **10**(1):57-63.
14. Suravajhala P, Kogelman LJ, Kadarmideen HN: **Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare**. *Genetics Selection Evolution* 2016, **48**(1):38.
15. Li J, Burmeister M: **Genetical genomics: combining genetics with gene expression analysis**. *Human Molecular Genetics* 2005, **14**(suppl 2):R163-R169.
16. Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast**. *Science* 2002, **296**(5568):752-755.
17. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C: **Mapping the genetic architecture of gene expression in human liver**. *PLoS Biol* 2008, **6**(5):e107.
18. Kadarmideen HN: **Systems Biology in Animal Production and Health**, vol. 2: Springer; 2016.
19. Alexandre PA, Kogelman LJ, Santana MH, Passarelli D, Pulz LH, Fantinato-Neto P, Silva PL, Leme PR, Strefezzi RF, Coutinho LL: **Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle**. *BMC genomics* 2015, **16**(1):1073.
20. Paradis F, Yue S, Grant J, Stothard P, Basarab J, Fitzsimmons C: **Transcriptomic analysis by RNA sequencing reveals that hepatic interferon-induced genes may be associated with feed efficiency in beef heifers**. *Journal of animal science* 2015, **93**(7):3331-3341.
21. Zhao X-M, Qin G: **Identifying biomarkers with differential analysis**. In: *Bioinformatics for Diagnosis, Prognosis and Treatment of Complex Diseases.* Springer; 2013: 17-31.
22. Lkhagvadorj S, Qu L, Cai W, Couture OP, Barb CR, Hausman GJ, Nettleton D, Anderson LL, Dekkers JC, Tuggle CK: **Gene expression profiling of the short-term adaptive response to acute caloric restriction in liver and adipose tissues of pigs differing in feed efficiency**. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 2010, **298**(2):R494-R507.
23. Tizioto PC, Coutinho LL, Oliveira PS, Cesar AS, Diniz WJ, Lima AO, Rocha MI, Decker JE, Schnabel RD, Mourão GB: **Gene expression differences in Longissimus muscle of Nelore steers genetically divergent for residual feed intake**. *Scientific Reports* 2016, **6**.
24. Armentano L, Weigel K: **Considerations for Improving Feed Efficiency in Dairy Cattle**. *Tuesday, October 22, 2013* 2013:37.

25. Tao H, Chang G, Xu T, Zhao H, Zhang K, Shen X: **Feeding a high concentrate diet down-regulates expression of ACACA, LPL and SCD and modifies milk composition in lactating goats**. *PloS one* 2015, **10**(6):e0130525.
26. Salleh M, Mazzoni G, Höglund J, Olijhoek D, Lund P, Løvendahl P, Kadarmideen H: **RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high-and low-residual feed intake in Nordic dairy cattle**. *BMC genomics* 2017, **18**(1):258.
27. Kogelman LJ, Cirera S, Zhernakova DV, Fredholm M, Franke L, Kadarmideen HN: **Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model**. *BMC medical genomics* 2014, **7**(1):57.
28. Kadarmideen HN, Watson-Haigh NS, Andronicos NM: **Systems biology of ovine intestinal parasite resistance: disease gene modules and biomarkers**. *Molecular Biosystems* 2011, **7**(1):235-246.
29. Kogelman LJ, Byrne K, Vuocolo T, Watson-Haigh NS, Kadarmideen HN, Kijas JW, Oddy HV, Gardner GE, Gondro C, Tellam RL: **Genetic architecture of gene expression in ovine skeletal muscle**. *BMC genomics* 2011, **12**(1):607.
30. Cogill SB, Wang L: **Co-expression network analysis of human lncRNAs and cancer genes**. *Cancer informatics* 2014(Suppl. 5):49.
31. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC bioinformatics* 2008, **9**(1):559.
32. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome biology* 2014, **15**(12):550.
33. [https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/faq.html]
34. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis**. *Statistical applications in genetics and molecular biology* 2005, **4**(1):1128.
35. Langfelder P, Zhang B, Horvath S: **Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R**. *Bioinformatics* 2008, **24**(5):719-720.
36. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J: **ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks**. *Bioinformatics* 2009, **25**(8):1091-1093.
37. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP: **STRING v10: protein–protein interaction networks, integrated over the tree of life**. *Nucleic acids research* 2014:gku1003.
38. Bauman DE, Currie WB: **Partitioning of nutrients during pregnancy and lactation: a review of mechanisms involving homeostasis and homeorhesis**. *Journal of dairy science* 1980, **63**(9):1514-1529.
39. Xi Y, Wu F, Zhao D, Yang Z, Li L, Han Z, Wang G: **Biological mechanisms related to differences in residual feed intake in dairy cows**. *animal* 2016, **10**(08):1311-1318.

40. Weber KL, Welly BT, Van Eenennaam AL, Young AE, Porto-Neto LR, Reverter A, Rincon G: **Identification of gene networks for residual feed intake in angus cattle using genomic prediction and RNA-seq**. *PloS one* 2016, **11**(3):e0152274.
41. Parhami F, Mody N, Gharavi N, Ballard AJ, Tintut Y, Demer LL: **Role of the cholesterol biosynthetic pathway in osteoblastic differentiation of marrow stromal cells**. *Journal of Bone and Mineral Research* 2002, **17**(11):1997-2003.
42. Karisa B, Moore S, Plastow G: **Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle**. *Animal Science Journal* 2014, **85**(4):374-387.
43. Graugnard DE, Piantoni P, Bionaz M, Berger LL, Faulkner DB, Loor JJ: **Adipogenic and energy metabolism gene networks in longissimus lumborum during rapid post-weaning growth in Angus and Angus× Simmental cattle fed high-starch or low-starch diets**. *BMC genomics* 2009, **10**(1):142.
44. Horodyska J, Oster M, Wimmers K, Mullen A, Lawlor P, Hamill R: **P3024 Transcriptome analysis of longissimus thoracis et lumborum from pigs divergent in residual feed intake**. *Journal of Animal Science* 2016, **94**(supplement4):63-64.
45. Nguyen TM, Sawyer JK, Kelley KL, Davis MA, Rudel LL: **Cholesterol esterification by ACAT2 is essential for efficient intestinal cholesterol absorption: evidence from thoracic lymph duct cannulation**. *Journal of lipid research* 2012, **53**(1):95-104.
46. Zhao Y, Hou Y, Liu F, Liu A, Jing L, Zhao C, Luan Y, Miao Y, Zhao S, Li X: **Transcriptome Analysis Reveals that Vitamin A Metabolism in the Liver Affects Feed Efficiency in Pigs**. *G3: Genes| Genomes| Genetics* 2016, **6**(11):3615-3624.
47. Hoogeveen RC, Reaves SK, Lei KY: **Copper deficiency increases hepatic apolipoprotein AI synthesis and secretion but does not alter hepatic total cellular apolipoprotein AI mRNA abundance in rats**. *The Journal of nutrition* 1995, **125**(12):2935.
48. Huster D, Purnat TD, Burkhead JL, Ralle M, Fiehn O, Stuckert F, Olson NE, Teupser D, Lutsenko S: **High copper selectively alters lipid metabolism and cell cycle machinery in the mouse model of Wilson disease**. *Journal of Biological Chemistry* 2007, **282**(11):8343-8355.
49. Richardson E, Herd R, Archer J, Arthur P: **Metabolic differences in Angus steers divergently selected for residual feed intake**. *Animal Production Science* 2004, **44**(5):441-452.
50. Arthur P, Archer J, Johnston D, Herd R, Richardson E, Parnell P: **Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle**. *Journal of animal science* 2001, **79**(11):2805-2811.
51. Schaefer EJ, Gleason JA, Dansinger ML: **Dietary fructose and glucose differentially affect lipid and glucose homeostasis**. *The Journal of nutrition* 2009, **139**(6):1257S-1262S.
52. Van Eerden E, Van Den Brand H, Parmentier H, De Jong M: **Phenotypic selection for residual feed intake and its effect on humoral immune responses in growing layer hens**. *Poultry science* 2004, **83**(9):1602-1609.
53. Mpetile Z: **Effect of divergent selection for residual feed intake on immune system of Yorkshire pigs**. 2014.

54. Ametaj B, Koenig K, Dunn S, Yang W, Zebeli Q, Beauchemin K: **Backgrounding and finishing diets are associated with inflammatory responses in feedlot steers**. *Journal of animal science* 2009, **87**(4):1314-1320.
55. Roberts A, Funston R, Mulliniks T, Petersen M, MacNeil M: **Feed efficiency-how should it be used for the cow herd?** 2011.
56. Schroder K, Hertzog PJ, Ravasi T, Hume DA: **Interferon-γ: an overview of signals, mechanisms and functions**. *Journal of leukocyte biology* 2004, **75**(2):163-189.
57. Gondret F, Vincent A, Houée-Bigot M, Siegel A, Lagarrigue S, Louveau I, Causeur D: **Molecular alterations induced by a high-fat high-fiber diet in porcine adipose tissues: variations according to the anatomical fat location**. *BMC genomics* 2016, **17**(1):120.
58. Reynolds J, Foote A, Freetly H, Oliver W, Lindholm-Perry A: **Relationships between inflammation-and immunity-related transcript abundance in the rumen and jejunum of beef steers with divergent average daily gain**. *Animal Genetics* 2017.
59. Jungermann K, Katz N: **Functional specialization of different hepatocyte populations**. *Physiological reviews* 1989, **69**(3):708-764.

Supplementary Materials (Paper 2)

Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency

Salleh, M.S.[1,2], Mazzoni, G. [2], Løvendahl, P.[3], and Kadarmideen, H.N.[2*]

[1] Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, DK-1870, Frederiksberg C, Denmark

[2] Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark

[3] Department of Molecular Biology and Genetics - Center for Quantitative Genetics and Genomics, Aarhus University, AU Foulum, DK-8830, Tjele, Denmark

* Corresponding author

SMS: surayams@sund.ku.dk

GM: gianmaz@bioinformatics.dtu.dk

PL: peter.lovendahl@mbg.au.dk

HNK: hajak@dtu.dk

ClueGO analysis output

Table 1 Significant (p<0.05) GO term and pathways in salmon module in Holstein

GO term function	PValue Corrected with Benjamini-Hochberg	Number of genes
cholesterol biosynthetic process	3.1E-04	27
GTPase binding	3.8E-04	12
organic hydroxy compound metabolic process	4.4E-04	16
collagen fibril organization	5.8E-04	10
microtubule organizing center	6.4E-04	22
Steroid biosynthesis	7.6E-04	34
cytoskeletal part	8.0E-04	30
astral microtubule organization	9.7E-04	50
Leukocyte transendothelial migration	1.4E-03	7
small GTPase mediated signal transduction	5.0E-03	14
oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	5.2E-03	3
Measles	1.1E-02	7
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	1.2E-02	5
sarcomere	1.4E-02	4
Protein oligomerization	1.9E-02	8
Glycerolipid metabolism	2.2E-02	3
lung alveolus development	2.2E-02	3
phagocytic vesicle	2.6E-02	3
cellular response to nitrogen compound	3.0E-02	8
solute:cation symporter activity	3.1E-02	3

Table 2 Significant (p<0.05) GO term and pathways in lightsteelblue1 module in Jersey

GO term function	PValue Corrected with Benjamini-Hochberg	Number of genes
positive regulation of interferon-gamma production	5.50E-10	11
lymphocyte differentiation	6,43E-09	13
side of membrane	2.47E-08	27
Natural killer cell mediated cytotoxicity	2,89E-08	15
external side of plasma membrane	6.97E-06	11
second-messenger-mediated signaling	6.10E-05	5
Cell adhesion molecules (CAMs)	4.36E-03	3

STRING 10 analysis for salmon module in Holstein and lightsteelblue1 in Jersey

Holstein

Table 3 Biological Process

#pathway ID	pathway description	observed gene count	FDR	matching proteins in your network (labels)
GO.0006695	cholesterol biosynthetic process	8	1.53E-09	*CYP51A1,DHCR7,FDFT1,HMGCR,IDI1,MVD,MVK,NSDHL*
GO.0006694	steroid biosynthetic process	9	8.49E-09	*CYP51A1,DHCR7,FDFT1,HMGCR,IDI1,LSS,MVD,MVK,NSDHL*
GO.0008610	lipid biosynthetic process	12	1.84E-06	*ACACA,ACLY,CYP51A1,DHCR7,FASN,FDFT1,GPAM,HMGCR,IDI1,LSS,MVK,NSDHL*
GO.0044283	small molecule biosynthetic process	10	0.000173	*ACACA,CYP51A1,DHCR7,FASN,FDFT1,HMGCR,IDI1,MVD,MVK,NSDHL*
GO.0006629	lipid metabolic process	13	0.00156	*ACACA,ACLY,CYP51A1,DHCR7,FASN,FDFT1,GPAM,HMGCR,IDI1,LSS,MVK,NCF1,NSDHL*
GO.0008299	isoprenoid biosynthetic process	4	0.00368	*FDFT1,HMGCR,IDI1,MVK*
GO.0022614	membrane to membrane docking	2	0.0468	*EZR,MSN*

Table 4 Cellular Component

#pathway ID	pathway description	observed gene count	FDR	matching proteins in your network (labels)
GO.0001931	uropod	2	0.0296	*EZR,MSN*
GO.0031254	cell trailing edge	2	0.0296	*EZR,MSN*
GO.0036064	ciliary basal body	4	0.0296	*CETN2,EZR,POC1A,PSEN2*
GO.0016324	apical plasma membrane	5	0.0365	*AMOTL2,CLDN1,EZR,MSN,PSEN2*
GO.0044444	cytoplasmic part	35	0.0418	*ACACA,ACLY,ACTR3,AMOTL2,BAIAP2,CETN2,CKB,COPG2,CTSK,CYP51A1,DHCR7,EIF2C2,FAM109B,FASN,FDFT1,GPAM,HMGCR,IDI1,IKBIP,LSS,MAP1S,MIA3,MOSPD1,MVD,MVK,NCF1,NSDHL,POC1A,RAB11A,RAB8B,SHOC2,SLC1A4,SLC25A37,SLC35A3,WDR44*

Table 5 KEGG pathways

#pathway ID	pathway description	observed gene count	FDR	matching proteins in your network (labels)
100	Steroid biosynthesis	9	2.85E-12	*CYP51A1,DHCR24,DHCR7,FDFT1,HSD17B7,LSS,MSMO1,NSDHL,SQLE*
900	Terpenoid backbone biosynthesis	6	1.35E-06	*ACAT2,HMGCR,HMGCS1,IDI1,MVD,MVK*
1100	Metabolic pathways	29	2.29E-05	*ACACA,ACAT2,ACLY,ACSS2,CKB,CSAD,CYP51A1,CYP7A1,DGKD,DHCR24,DHCR7,ELOVL6,FASN,FDFT1,GPAM,HMGCR,HMGCS1,HSD17B7,IDI1,IDUA,LSS,MBOAT2,MSMO1,MVD,MVK,NSDHL,PHOSPHO2,RDH11,SQLE*
1212	Fatty acid metabolism	5	0.00386	*ACACA,ACAT2,ELOVL6,FADS1,FASN*
4670	Leukocyte transendothelial migration	7	0.00386	*CLDN1,EZR,GNAI1,MSN,NCF1,PIK3CG,RAPGEF4*

Jersey

Table 6 Biological Process

#pathway ID	pathway description	observed gene count	FDR	matching proteins in your network (labels)
GO.0050870	positive regulation of T cell activation	4	1.03E-02	*CCL5,CD3E,CD5,SASH3*
GO.0002684	positive regulation of immune system process	6	1.16E-02	*Bt.87330,CCL5,CD3E,CD5,CXCL9,SASH3*
GO.0070098	chemokine-mediated signaling pathway	3	1.62E-02	*CCL5,CCR5,CXCL9*
GO.0002376	immune system process	7	0.0203	*CCL5,CCR5,CD3D,CD3E,CD5,PSMB8,PSMB9*
GO.0032753	positive regulation of interleukin-4 production	2	0.0203	*CD3E,SASH3*
GO.0042102	positive regulation of T cell proliferation	3	0.0203	*CCL5,CD3E,SASH3*
GO.0045061	thymic T cell selection	2	0.0203	*CD3D,CD3E*
GO.0016337	single organismal cell-cell adhesion	4	0.024	*CCL5,CD3D,CD3E,ICAM3*
GO.0031295	T cell costimulation	2	0.0269	*CD3E,CD5*
GO.0007155	cell adhesion	5	0.0476	*CCL5,CD3D,CD3E,CD96,ICAM3*

Table 7 Cellular Component

#pathway ID	pathway description	observed gene count	FDR	matching proteins in your network (labels)
GO.0009897	external side of plasma membrane	4	3.51E-03	*CCR5,CD3E,CD5,CXCL9*
GO.0042105	alpha-beta T cell receptor complex	2	8.53E-03	*CD3D,CD3E*
GO.1990111	spermatoproteasome complex	2	1.02E-02	*PSMB8,PSMB9*

Table 8 KEGG Pathway

#pathway ID	pathway description	observed gene count	FDR	matching proteins in your network (labels)
4650	Natural killer cell mediated cytotoxicity	7	1.97E-06	*CD244,KLRK1,LAT,LCK,LCP2,PRF1,ZAP70*
4660	T cell receptor signaling pathway	7	1.97E-06	*CARD11,CD3D,CD3E,LAT,LCK,LCP2,ZAP70*

5340	Primary immunodeficiency	4	2.21E-04	*CD3D,CD3E,LCK,ZAP70*
4064	NF-kappa B signaling pathway	5	0.000279	*CARD11,LAT,LCK,PTGS2,ZAP70*
4640	Hematopoietic cell lineage	4	0.00321	*CD2,CD3D,CD3E,CD5*
5166	HTLV-I infection	6	0.00321	*CD3D,CD3E,HLA-DOA,JAK3,JSP.1,LCK*
4062	Chemokine signaling pathway	5	0.00428	*CCL5,CCR5,CXCL9,CXCR6,JAK3*
5332	Graft-versus-host disease	3	0.00443	*HLA-DOA,JSP.1,PRF1*
5330	Allograft rejection	3	0.00583	*HLA-DOA,JSP.1,PRF1*
4060	Cytokine-cytokine receptor interaction	5	0.00699	*CCL5,CCR5,CD27,CXCL9,CXCR6*
4940	Type I diabetes mellitus	3	0.00699	*HLA-DOA,JSP.1,PRF1*
5320	Autoimmune thyroid disease	3	0.00699	*HLA-DOA,JSP.1,PRF1*
4514	Cell adhesion molecules (CAMs)	4	0.0122	*CD2,HLA-DOA,ICAM3,JSP.1*
5416	Viral myocarditis	3	0.0122	*HLA-DOA,JSP.1,PRF1*
5203	Viral carcinogenesis	4	0.0394	*Bt.87330,CCR5,JAK3,JSP.1*

Upstream regulator from IPA® analysis

Table 9 Upstream regulators for Holstein

Upstream Regulator	Expr Log Ratio	Molecule Type	Predicted Activation State	Bias-corrected z-score	p-value of overlap	Target molecules in dataset	Mechanistic Network
ATP7B	0,162	transporter	Activated	2,747	1,19E-20	ACLY,CYP51A1,CYP7A1,ELOVL6,FASN,FDFT1,FDPS,HMGCR,HMGCS1,IDI1,LSS,MSMO1,SQLE,SREBF2	
POR	0,113	enzyme	Inhibited	-2,653	2,66E-19	ACAT2,ACLY,CSAD,CYB5B,CYP51A1,CYP7A1,DHCR24,DHCR7,ELOVL6,FDFT1,FDPS,HMGCR,HMGCS1,IDI1,LSS,MSMO1,MVD,MVK,NSDHL,SQLE,SREBF2	
cholesterol		chemical - endogenous mammalian	Inhibited	-3,898	6,09E-18	ACLY,ACSS2,CTSK,CYB5B,CYP7A1,DHCR7,FADS1,FASN,FCGR2B,FDFT1,FDPS,GNAI1,HMGCR,HMGCS1,IDI1,LYZ,MSMO1,NSDHL,RDH11,SQLE,SREBF2	53 (12)

Table 10 Upstream regulators for Jersey

Upstream Regulator	Expr Log Ratio	Molecule Type	Predicted Activation State	Bias-corrected z-score	p-value of overlap	Target molecules in dataset	Mechanistic Network
IFNG		cytokine	Inhibited	-2,068	1,43E-06	CCL5,CCR5,CD2,CXCL9,GBP5,HLA-B,HLA-DOA,JAK3,LAMP3,LCP2,PSMB8,PSMB9,PTGS2,UBD	20 (12)
IL10RA	-0,326	transmembrane receptor	Activated	2,526	1,04E-04	CCL5,GBP5,Klrk1,PSMB8,PSMB9,TRPM2	12 (4)
NKX2-3		transcription regulator	Activated	2,035	1,96E-03	HLA-B,PSMB8,PSMB9,PTGS2	
dexamethasone		chemical drug	Activated	2,67	5,26E-02	CCL5,CD3D,CD3E,ICAM3,JAK3,PFKFB3,PTGS2,UBD	

Disease and functions for most significant modules

Table 11 Diseases and functions in Salmon module in Holstein cows

Categories	Diseases or Functions Annotation	p-Value	Predicted Activation State	Activation z-score	Molecules	# Molecules
Lipid Metabolism, Small Molecule Biochemistry, Vitamin and Mineral Metabolism	synthesis of cholesterol	3,66E-17	Increased	2.321	ACLY,CYP51A1,CYP7A1,DHCR24,DHCR7,FDFT1,FDPS,HMGCR,HSD17B7,HTT,IDI1,LSS,MVK,SREBF2	14
Lipid Metabolism, Small Molecule Biochemistry, Vitamin and Mineral Metabolism	synthesis of sterol	6,08E-17	Increased	2.573	ACAT2,ACLY,CYP51A1,CYP7A1,DHCR24,DHCR7,FDFT1,FDPS,HMGCR,HSD17B7,HTT,IDI1,LSS,MVK,SREBF2	15
Lipid Metabolism, Small Molecule Biochemistry, Vitamin and Mineral Metabolism	metabolism of cholesterol	2,67E-16	Increased	2.405	ACLY,CYP51A1,CYP7A1,DHCR24,DHCR7,FDFT1,FDPS,HMGCR,HSD17B7,HTT,IDI1,LSS,MVK,NSDHL,SQLE,SREBF2	16
Lipid Metabolism, Small Molecule Biochemistry	metabolism of membrane lipid derivative	1,21E-11	Increased	2.922	ACLY,CYP51A1,CYP7A1,DHCR24,DHCR7,FADS1,FASN,FCGR2B,FDFT1,FDPS,GPAM,HMGCR,HSD17B7,HTT,IDI1,LSS,MVK,NSDHL,PIK3CG,RHOQ,SQLE,SREBF2	22
Lipid Metabolism, Small Molecule Biochemistry	synthesis of lipid	3,92E-10	Increased	3.616	ACACA,ACAT2,ACLY,ACSS2,BMP4,CYP51A1,CYP7A1,DGKD,DHCR24,DHCR7,ELOVL6,FADS1,FASN,FCGR2B,FDFT1,FDPS,GPAM,HMGCR,HSD17B7,HTT,IDI1,LSS,MVD,MVK,NCF1,PIK3CG,RHOQ,SREBF2	28
Lipid Metabolism, Nucleic	conversion of	1,47E-08	Increased	2.200	ACACA,ACLY,ACSS2,FASN,H	5

Acid Metabolism, Small Molecule Biochemistry	acyl-coenzyme A				MGCR	
Lipid Metabolism, Small Molecule Biochemistry	conversion of lipid	5,51E-07	Increased	2.967	ACACA,ACLY,ACSS2,CYB5B,DHCR24,FADS1,FASN,FCGR2B,HMGCR,HTT	10
Lipid Metabolism, Small Molecule Biochemistry	conversion of fatty acid	1,30E-06	Increased	2.414	ACACA,ACLY,ACSS2,FADS1,FASN,HMGCR	6
Cellular Assembly and Organization, Cellular Function and Maintenance	organization of cytoplasm	1,07E-05	Increased	4.531	ACACA,ACTR3,ANKRD27,BAIAP2,BICDL1,BMP4,CEP120,CETN2,CGN,CHD3,DNAJC13,EVL,EZR,FASN,HTT,IDUA,KIF3B,LCK,LCP1,MAP1S,MAPKAPK5,MSN,PRKCH,PSEN2,PTPRE,RAB11A,RALBP1,RAPGEF4,RHOQ,RYR1,SPTAN1,STK26,WIPF1,YME1L1	34
Cellular Assembly and Organization, Cellular Function and Maintenance	organization of cytoskeleton	2,47E-05	Increased	4.507	ACACA,ACTR3,ANKRD27,BAIAP2,BICDL1,BMP4,CEP120,CETN2,CGN,CHD3,EVL,EZR,FASN,HTT,KIF3B,LCK,LCP1,MAP1S,MAPKAPK5,MSN,PRKCH,PSEN2,PTPRE,RAB11A,RALBP1,RAPGEF4,RHOQ,RYR1,SPTAN1,STK26,WIPF1	31
Infectious Diseases	Viral Infection	4,76E-05	Increased	3.270	ACTR3,AGO2,CDC40,CYB5B,CYP51A1,FASN,FCGR2B,FDFT1,FDPS,HLA-DOA,HMGCR,HMGCS1,IFIH1,IGHMBP2,KIAA0922,LCK,MAP1S,MYOF,NCF1,PIK3CG,PRKCH,PSEN2,RAB11A,RAB8B,SLU7,SPRY2,SPTAN1,SREBF2,TMC8,TRIM5,TUBB2A,WIPF1,XK,ZNF791	34

Cell Morphology, Cellular Assembly and Organization, Cellular Function and Maintenance	formation of cellular protrusions	6,81E-05	Increased	4.031	ACACA,ACTR3,ANKRD27,BAIAP2,BICDL1,CEP120,CETN2,EZR,FASN,HTT,KIF3B,LCP1,MAP1S,MSN,PSEN2,PTPRE,RAB11A,RAPGEF4,RHOQ,RYR1,STK26,WIPF1	22
Cellular Assembly and Organization, Cellular Function and Maintenance	microtubule dynamics	1,23E-04	Increased	4.408	ACACA,ACTR3,ANKRD27,BAIAP2,BICDL1,BMP4,CEP120,CETN2,CGN,EVL,EZR,FASN,HTT,KIF3B,LCP1,MAP1S,MSN,PRKCH,PSEN2,PTPRE,RAB11A,RAPGEF4,RHOQ,RYR1,STK26,WIPF1	26
Infectious Diseases	infection by lentivirus	1,25E-04	Increased	2.458	ACTR3,CDC40,CYP51A1,FCGR2B,FDFT1,FDPS,HLA-DOA,HMGCR,HMGCS1,IGHMBP2,KIAA0922,MYOF,PRKCH,PSEN2,SLU7,SPTAN1,TRIM5,TUBB2A,ZNF791	19
Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry	concentration of sterol	1,40E-04	Increased	2.357	ACAT2,CYP7A1,DHCR24,DHCR7,FDFT1,GPAM,HMGCR,HTT,KLB,PSEN2,SREBF2	11
Organismal Survival	organismal death	1,46E-04	Decreased	-4.455	ACACA,ACLY,AGO2,BMP4,CETN2,CHEK2,CLDN1,COL4A1,CSAD,CYP51A1,CYP7A1,DGKD,DHCR7,E2F3,ELK3,FASN,FCGR2B,FDFT1,GNAI1,HMGCR,HS6ST1,HSD17B7,HTT,IDUA,IFIH1,KIF3B,LCP1,LPAR6,LYZ,MAPKAPK5,MIA3,MSN,NCF1,PIK3CG,PRKCH,PSEN2,RAB11A,RAB8B,RYR1,SHOC2,SPRY2,SREBF2,TCEA1,WIPF1	44

Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry	concentration of lipid	2,07E-04	Increased	3.143	ACACA,ACAT2,ACLY,CYP7A1, DHCR24,DHCR7,E2F3,FASN, FCGR2B,FDFT1,GNAI1,GPA M,HMGCR,HS6ST1,HTT,IGH MBP2,KLB,PIK3CG,PSEN2,SR EBF2	20
Cell Morphology, Cellular Assembly and Organization, Cellular Function and Maintenance	reorganization of cytoskeleton	2,59E-04	Increased	2.213	BAIAP2,EZR,LCK,LCP1,MAPK APK5,MSN,RHOQ,SPTAN1	8
Infectious Diseases	HIV infection	3,53E-04	Increased	2.776	ACTR3,CDC40,CYP51A1,FCG R2B,FDFT1,FDPS,HLA-DOA,HMGCR,HMGCS1,IGH MBP2,KIAA0922,MYOF,PRK CH,PSEN2,SLU7,SPTAN1,TU BB2A,ZNF791	18
Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry	concentration of cholesterol	3,88E-04	Increased	2.263	CYP7A1,DHCR24,DHCR7,FDF T1,GPAM,HMGCR,HTT,KLB,P SEN2,SREBF2	10
Cellular Growth and Proliferation, Tissue Development	generation of cells	1,02E-03	Increased	2.328	ACTR3,AGO2,ANKRD27,BAIA P2,BICDL1,BMP4,CD3D,CGN, COL4A1,DDIAS,DHCR24,EZR, FCGR2B,Fmnl1,HLA-DOA,HS6ST1,HTT,KIF3B,LCK, LOXL2,MAP1S,MIA3,MSN,M YOF,PIK3CG,PRKCH,PSEN2,P TPRE,RAB11A,RALBP1,RAPG EF4,RHOQ,RYR1,SFMBT1,TC F7,WIPF1	36
Developmental Disorder	hypoplasia of organ	1,21E-03	Decreased	-3.399	BMP4,CYP51A1,CYTIP,E2F3, HS6ST1,HSD17B7,LCK,PIK3C G,PRKCH,SPRY2,TCEA1,WIPF	12

Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry	quantity of steroid	1,92E-03	Increased	3.157	1 ACAT2,CYP7A1,DHCR24,DHCR7,FDFT1,GPAM,HMGCR,HTT,IGHMBP2,KLB,PSEN2,SREBF2	12
Lipid Metabolism, Small Molecule Biochemistry	fatty acid metabolism	2,05E-03	Increased	2.423	ACACA,ACLY,ACSS2,CYP7A1,ELOVL6,FADS1,FASN,FCGR2B,GPAM,HTT,LSS,MSMO1,MSN,NCF1	14
Cellular Growth and Proliferation	proliferation of cells	2,23E-03	Increased	2.733	ACACA,ACLY,AGO2,BAIAP2,BMP4,CEP120,CHEK2,CLDN1,COL4A1,CSNK1G3,CYTIP,DGKD,DHCR24,DHCR7,E2F3,EZR,FADS1,FASN,FCGR2B,FDFT1,GNAI1,GPAM,HMGCR,HS6ST1,HTT,KLB,LCK,LCP1,LOXL2,LRRFIP1,LYZ,MMP11,MVD,MYOF,NCF1,PIK3CG,PIK3IP1,PPP1CB,PRKCH,PSEN2,PTPRE,RAB11A,RAB8B,RALBP1,RHOQ,SKAP2,SPRY2,SPTAN1,STK26,TCF7,THAP12,TMC8,TUBB2A,USP36,WIPF1,YME1L1	56
Infectious Diseases	infection of cells	2,43E-03	Increased	2.509	ACTR3,CDC40,CYB5B,FCGR2B,HLA-DOA,HMGCR,HMGCS1,IGHMBP2,KIAA0922,MYOF,PRKCH,PSEN2,RAB8B,SLU7,SPTAN1,TRIM5,ZNF791	17
Developmental Disorder	dysgenesis	2,62E-03	Decreased	-3.393	BMP4,CYP51A1,CYTIP,E2F3,HS6ST1,HSD17B7,LCK,PIK3CG,PRKCH,SLC1A4,SPRY2,TCEA1,WIPF1	13

Lipid Metabolism, Small Molecule Biochemistry	metabolism of phospholipid	4,89E-03	Increased	2.142	CYP7A1,FADS1,FASN,FCGR2B,GPAM,PIK3CG,RHOQ ACACA,ACLY,AGO2,BMP4,CDC40,CHEK2,E2F3,EZR,FASN,HTT,IFIH1,LCK,LYZ,PIK3CG,PIK3IP1,PPP1CB,PRKCH,PSEN2,PTPRE,RAB11A,RAPGEF4,SREBF2,TCF7,USP36	7
Cell Death and Survival	cell viability	4,95E-03	Increased	3.511		24

Table 12 Diseases and functions in Magenta module in Holstein cows

Categories	Diseases or Functions Annotation	p-Value	Predicted Activation State	Activation z-score	Molecules	# Molecules
Cell Cycle	senescence of cells	2,14E-04	Decreased	-2.078	BHLHE40,BRCA1,DUSP1,FANCD2,GADD45A,MAPK9,NAMPT,NFE2L2,NRAS,PAX8,SRF	11
Lipid Metabolism, Small Molecule Biochemistry	synthesis of fatty acid	3,52E-04	Decreased	-2.066	ABAT,ACADL,ACSL1,APOA4,APOA5,BRCA1,CD14,ELOVL2,FOXA1,HNF1A,MAPK9,MID1IP1	12
Carbohydrate Metabolism	quantity of monosaccharide	8,66E-04	Increased	2.359	ACSL1,ADIPOR2,ADM,ALDH1A1,APOA4,C1QTNF12,DGAT2,FOXA1,HNF1A,MADD,NRAS,RGS16	12
Carbohydrate Metabolism, Molecular Transport, Small Molecule Biochemistry	concentration of D-glucose	2,41E-03	Increased	2.359	ACSL1,ADIPOR2,ADM,ALDH1A1,APOA4,C1QTNF12,DGAT2,FOXA1,HNF1A,MADD,RGS16	11
Carbohydrate Metabolism	uptake of carbohydrate	2,82E-03	Decreased	-2.343	ACSL1,ADM,APOA1,C1QTNF12,CD14,CRAT,HNF1A,MAPK9,NR1I2,SRF	10
Molecular Transport	export of molecule	6,68E-03	Decreased	-2.124	ACSL1,APOA1,APOA4,APOA5,CD14,CRAT,NR1I2,SLC22A7,SOAT2	9

Table 13 Diseases and functions in Lightsteelblue1 module in Jersey cows

Categories	Diseases or Functions Annotation	p-Value	Predicted Activation State	Activation z-score	Molecules	# Molecules
Cell-To-Cell Signaling and Interaction, Hematological System Development and Function, Immune Cell Trafficking, Inflammatory Response	activation of leukocytes	4.66E-21	Decreased	-2.61	CARD11,CCL5,CCR5,CD2,CD27,CD3D,CD3E,CD5,CXCL9,GIMAP1-GIMAP5,HLA-DOA,ICAM3,JAK3,Klre1,Klrk1,LAT,LCK,LCP2,PRF1,PSMB8,PSMB9,PTGS2,SASH3,TRPM2,ZAP70	25
Hematological System Development and Function, Lymphoid Tissue Structure and Development, Tissue Morphology	quantity of T lymphocytes	2.03E-14	Decreased	-3.43	CARD11,CCL5,CCR5,CD27,CD3D,CD3E,CD5,GIMAP1-GIMAP5,JAK3,Klrk1,LAT,LCK,LCP2,PRF1,PSMB8,PSMB9,SASH3,ZAP70	18
Cellular Development, Hematological System Development and Function, Hematopoiesis, Lymphoid Tissue Structure and Development	differentiation of leukocytes	5.18E-14	Decreased	-2.07	CARD11,CCL5,CD2,CD27,CD3D,CD3E,GIMAP1,GIMAP1-GIMAP5,HLA-DOA,JAK3,Klrk1,LAT,LCK,LCP2,PTGS2,SASH3,TRPM2,UBD,ZAP70	19
Hematological System Development and Function, Tissue Morphology	quantity of leukocytes	3.06E-12	Decreased	-3.44	CARD11,CCL5,CCR5,CD27,CD3D,CD3E,CD5,CXCR6,GIMAP1-GIMAP5,JAK3,Klrk1,LAT,LCK,LCP2,PRF1,PSMB8,PSMB9,PTGS2,SASH3,ZAP70	20
Cell Death and Survival, Cellular Compromise	cytotoxicity of lymphocytes	9.59E-12	Decreased	-2.22	CARD11,CCL5,CD2,CD27,CD5,CD96,Klrk1,LAT,LCK,PRF1	10
Cellular	Lymphocyte	3.41E-	Decreas	-2.04	CCL5,CCR5,CD2,CD3E	12

Movement, Hematological System Development and Function, Immune Cell Trafficking	migration	10	ed		,CXCL9,JAK3,LAT,LCK, LCP2,PRF1,PTGS2,ZA P70	
Developmental Disorder, Organismal Injury and Abnormalities	hypoplasia of lymphatic system	5.04E-10	Increase d	2.92	CD3E,GIMAP1-GIMAP5,JAK3,Klrk1,L AT,LCK,LCP2,PRF1,SA SH3	9
Cell-To-Cell Signaling and Interaction, Hematological System Development and Function	interaction of T lymphocytes	3.64E-09	Decreas ed	-2.20	CCL5,CCR5,CD2,CXCL 9,ICAM3,LCK,LCP2	7
Cell Signaling, Molecular Transport, Vitamin and Mineral Metabolism	mobilization of Ca2+	4.41E-09	Decreas ed	-2.33	CCL5,CCR5,CD2,CD3E ,CD5,CXCL9,CXCR6,L AT,LCK,LCP2,ZAP70	11
Developmental Disorder, Immunological Disease, Organismal Injury and Abnormalities	hypoplasia of lymphoid organ	9.40E-09	Increase d	2.75	CD3E,GIMAP1-GIMAP5,JAK3,Klrk1,L AT,LCK,LCP2,PRF1	8
Cell-To-Cell Signaling and Interaction, Hematological System Development and Function, Immune Cell Trafficking, Inflammatory Response	activation of natural killer cells	4.84E-08	Decreas ed	-2.19	CARD11,CCL5,CD2,C D27,Klre1,Klrk1,PRF1	7
Cell-To-Cell Signaling and Interaction, Hematological System	adhesion of immune cells	7.06E-08	Decreas ed	-2.96	CCL5,CCR5,CD2,CXCL 9,ICAM3,JAK3,LCK,LC P2,PTGS2,ZAP70	10

Development and Function, Immune Cell Trafficking						
Cellular Development	differentiation of cells	7.89E-08	Decreased	-2.27	ARHGEF3,CARD11,CCL5,CCR5,CD2,CD27,CD3D,CD3E,GIMAP1,GIMAP1-GIMAP5,HLA-DOA,HOPX,JAK3,Klrk1,LAT,LCK,LCP2,PSMB8,PTGS2,SASH3,TRPM2,UBD,Wfdc21,ZAP70	24
Cell-To-Cell Signaling and Interaction	aggregation of cells	5.99E-07	Decreased	-2.34	CCL5,CD2,CXCL9,ICAM3,LAT,LCK,LCP2,PTGS2,ZAP70	9
Cellular Movement, Hematological System Development and Function, Immune Cell Trafficking	cell movement of leukocytes	8.45E-07	Decreased	-2.42	CCL5,CCR5,CD2,CD3E,CXCL9,JAK3,LAT,LCK,LCP2,PRF1,PTGS2,TRPM2,ZAP70	13
Free Radical Scavenging	synthesis of reactive oxygen species	2.00E-06	Decreased	-2.12	CCL5,CCR5,CXCL9,LAT,LCK,LCP2,PRF1,PTGS2,TRPM2,ZAP70	10
Cell-To-Cell Signaling and Interaction, Hematological System Development and Function, Immune Cell Trafficking	adhesion of mononuclear leukocytes	2.39E-06	Decreased	-2.43	CCL5,CD2,CXCL9,JAK3,LCP2,ZAP70	6
Developmental Disorder	Hypoplasia	3.23E-06	Increased	3.08	CD3E,GIMAP1-GIMAP5,JAK3,Klrk1,LAT,LCK,LCP2,PRF1,PTGS2,SASH3	10
Cell-To-Cell Signaling and Interaction, Cell-mediated Immune Response, Cellular Movement,	adhesion of T lymphocytes	4.66E-06	Decreased	-2.21	CD2,CXCL9,JAK3,LCP2,ZAP70	5

Hematological System Development and Function, Immune Cell Trafficking						
Developmental Disorder, Immunological Disease, Organismal Injury and Abnormalities	hypoplasia of thymus gland	7.31E-06	Increased	2.17	CD3E,JAK3,LAT,LCK,PRF1	5
Developmental Disorder	hypoplasia of organ	8.08E-06	Increased	2.92	CD3E,GIMAP1-GIMAP5,JAK3,Klrk1,LAT,LCK,LCP2,PRF1,PTGS2	9
Cellular Movement	homing of cells	7.61E-05	Decreased	-2.35	CCL5,CCR5,CXCL9,CXCR6,JAK3,LAT,LCK,PTGS2,TRPM2	9
Cellular Growth and Proliferation	proliferation of cells	1.98E-04	Decreased	-2.42	CARD11,CCL5,CCR5,CD2,CD27,CD3E,CD5,CXCL9,DNAJA1,GPR174,HOPX,ICAM3,JAK3,Klrk1,LAT,LCK,LCP2,PFKFB3,PRF1,PTGS2,RASSF4,SASH3,TRPM2,UBD,ZAP70	25
Cellular Movement	cell movement	6.20E-04	Decreased	-2.36	ABI3,ARHGEF28,CCL5,CCR5,CD2,CD3E,CXCL9,CXCR6,DNAJA1,JAK3,LAT,LCK,LCP2,PRF1,PTGS2,TRPM2,ZAP70	17
Cellular Movement	migration of cells	1.45E-03	Decreased	-2.14	ABI3,ARHGEF28,CCL5,CCR5,CD2,CD3E,CXCL9,JAK3,LAT,LCK,LCP2,PRF1,PTGS2,TRPM2,ZAP70	15
Cellular Movement	chemotaxis	1.64E-03	Decreased	-2.13	CCL5,CCR5,CXCL9,LAT,LCK,PTGS2,TRPM2	7

Table 14 Diseases and functions in Violet module in Jersey cows

Categories	Diseases or Functions Annotation	p-Value	Predicted Activation State	Activation z-score	Molecules	# Molecules
Gastrointestinal Disease, Organismal Injury and Abnormalities	abnormality of large intestine	3,53E-03	Decreased	-2	CBLB,FAAH,MERTK,MPC1,PLD1,PON1,TP53,ZNF784	8
Gastrointestinal Disease, Inflammatory Disease, Inflammatory Response, Organismal Injury and Abnormalities	colitis	7,08E-03	Decreased	-2	FAAH,MERTK,MPC1,PLD1,PON1,TP53,ZNF784	7

Manuscript 3: Identification of expression QTLs targeting candidate genes for residual feed intake in Danish dairy cattle using systems genomics

ISSN: 2378-3648

Salleh et al. J Genet Genome Res 2018, 5:035
DOI: 10.23937/2378-3648/1410035
Volume 5 | Issue 1
Open Access

Journal of
Genetics and Genome Research

RESEARCH ARTICLE

Identification of Expression QTLs Targeting Candidate Genes for Residual Feed Intake in Dairy Cattle Using Systems Genomics

Salleh MS[1,2], Mazzoni G[2], Nielsen MO[1], Løvendahl P[3] and Kadarmideen HN[2,4*]

[1]*Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark*

[2]*Department of Bio and Health Informatics, Technical University of Denmark, Lyngby, Denmark*

[3]*Department of Molecular Biology and Genetics-Center for Quantitative Genetics and Genomics, Aarhus University, AU Foulum, Tjele, Denmark*

[4]*Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark*

***Corresponding author:** *Kadarmideen HN, Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark, E-mail: hajak@dtu.dk*

Abstract

Background: Residual feed intake (RFI) is the difference between actual and predicted feed intake and an important factor determining feed efficiency (FE). Recently, 170 candidate genes were associated with RFI, but no expression quantitative trait loci (eQTL) mapping has hitherto been performed on FE related genes in dairy cows. In this study, an integrative systems genetics approach was applied to map eQTLs in Holstein and Jersey cows fed two different diets to improve identification of candidate genes for FE.

Methods: Liver RNA-seq transcriptomics data from nine Holstein and ten Jersey cows that had been fed control (C) or high concentrate (HC) diets were integrated with genomic data (from 777k BovineHD Illumina BeadChip) by using the Matrix eQTL R package. A total of 170 previously identified candidate genes for FE (89 differentially expressed genes (DEGs) between high and low RFI groups and 81 hub genes (HG) in a group of co-expressed genes) were used in the data integration analysis.

Results: From the 241,542 SNPs used in the analysis, we identified 20 significant (FDR < 0.05) local-eQTLs targeting seven candidate genes and 16 significant (FDR < 0.05) local-eQTLs targeting five candidate genes related to RFI for the C and HC diet group analysis, respectively, in a breed-specific way.

Conclusions: Interestingly, Holstein and Jersey cows appear to rely on different strategies (lipid and cholesterol metabolism versus immune and inflammatory function) to achieve low RFI. The eQTLs overlapped with QTLs previously associated with FE trait (e.g. dry matter intake, longevity, body weight gain and net merit). The eQTLs and biological pathways identified in this study improve our understanding of the complex biological and genetic mechanisms that determine FE traits in dairy cattle. The identified eQTLs/genetic variants can potentially be used in new genomic selection methods that include biological/functional information on SNPs.

Keywords

eQTL, RNA-seq, Genotype, Data integration, Systems genomics, Feed efficiency, Residual feed intake

Abbreviations

ANOVA: Analysis of Variance; C: Low Concentrate (Control); DCRC: Danish Cattle Research Centre; DEGs: Differentially Expressed Genes; EDTA: Ethylenediaminetetraacetic Acid; eQTL: Expression Quantitative Trait Loci; FDR: False Discovery Rate; FE: Feed Efficiency; HC: High Concentrate; HG: Hub Genes; Mb: Mega Base; QTL: Quantitative Trait Loci; RFI: Residual Feed Intake; RNA: Ribonucleic Acid; RNA-seq: RNA Sequencing; SNPs: Single Nucleotide Polymorphisms; WGCNA: Weighted Gene Co-expression Analysis

Background

Feed intake and the conversion of absorbed nutrients into milk components are major determinants of feed efficiency (FE) in dairy cattle and hence production economics. FE is a complex trait that is influenced by several genetic and environmental factors, which in an

Citation: Salleh MS, Mazzoni G, Nielsen MO, Løvendahl P, Kadarmideen HN (2018) Identification of Expression QTLs Targeting Candidate Genes for Residual Feed Intake in Dairy Cattle Using Systems Genomics. J Genet Genome Res 5:035. doi.org/10.23937/2378-3648/1410035

Accepted: July 14, 2018: **Published:** July 16, 2018

interactive way control feed intake, nutrient partitioning and metabolic adaptation to lactation in different body tissues as well as milk synthesis and immune function. In dairy cattle, the use of FE for breeding purposes is quite complicated, since recording of individual feed intake is difficult in group fed cows. It is therefore desirable to be able to predict the genetic contributions to this trait to be able to select the most feed efficient cows for breeding purposes.

To date, transcriptomics has given precise and reliable results that identify candidate genes related to phenotypes of interest [1]. Although gene expressions associated with FE related genes have been studied for a long time, also in cattle [2-4], genetic markers are more easily accessible and not affected by environmental factors in contrast to gene expression data.

However, in some cows, the actual feed intake deviate from the predicted by their genetic heritage, even when they are exposed to similar environmental conditions. The term residual feed intake (RFI) describes this deviation and is calculated as the difference between the actual measured and the predicted feed intake of the cow [5]. Among groups of high and low RFI cattle, we have recently identified several candidate genes that predict the RFI in Danish dairy cattle [6].

Therefore, in this present study we focused on genetic markers for RFI in an attempt to improve the prediction of genetic merit for FE, which is needed to be able to use this type of determinants in practice.

Integration of transcriptomics and genomics data can be used to identify potential causal genetic variants that affect particular phenotypes. This approach is known as Genetical Genomics or Integrative Genomics [7]. The identified regions are called expression Quantitative Trait Loci (eQTL). In other words, an eQTL is a region in a particular locus that influences or controls the differences of expressions of causal genes [8-11]. The expression profile is an intermediate biological space between the phenotype and the genome. Therefore, eQTL analysis can identify interesting genetic variants even with a low sample size [12,13]. The identification of genomic regions influencing the expressions of the candidate genes could give a better perspective to use the information in animal selection as well as provide a better explanation about the way genomic regions control traits of interest.

A few studies have been conducted to identify genomic regions determining FE traits in beef cattle, chicken and other livestock species [14-17]. However, no eQTL mapping has hitherto been performed on FE related genes in dairy cows.

In this study, we performed an eQTL mapping analysis on candidate genes for the RFI trait. The hypothesis of the integrative genomics analysis is that SNPs associated with the expression of candidate genes are involved or in linkage with genomic regions regulating their expression. Therefore, the objective of this study was to identify eQTL regions together with their functional annotations associated with the RFI trait in two breeds of dairy cattle (Danish Holstein and Danish Jersey) fed two different diets and to present an eQTL mapping of candidate genes for RFI using matrix eQTL analysis, as well as characterize the SNPs by comparing our findings with previously annotated QTLs. The eQTL identified in this study could be important candidate genetic markers defining actual FE in dairy cattle, and our study suggests that there are differential traits relating to RFI in Danish Holsteins as compared to Jerseys.

Materials and Methods

Experimental animals, RFI characteristics and experimental design

The present study is based on biological samples obtained from nine Holstein and ten Jersey cows, housed at the Danish Cattle Research Centre (DCRC), Aarhus University, Denmark. The animals were part of a larg-

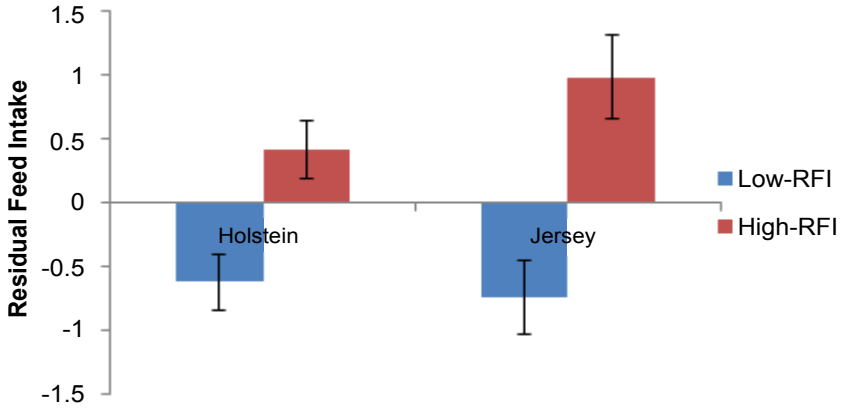


Figure 1: Mean ± SE of Residual Feed Intake (RFI) value for the Holstein and Jersey cows used in the present experiment [6].

er experiment, where FE was determined in 200 dairy cows distributed on the two breeds [18]. The details about the animal's background and the overall experimental design of the larger trial can be found in Salleh, et al. and Li, et al. [6,18].

The experimental cows used in the present study were selected based on individually recorded RFI of cows from the larger study. A total of four Holstein cows with very high and five with very low RFI, and five Jersey cows with very high and five with very low RFI were selected, and their deviation from the average recorded RFI is shown in Figure 1. The experimental cows underwent two periods of feeding trials low concentrate (control (C)) and high concentrate (HC) diet. The two dietary exposures were separated by a conditioning period of 14 - 26 days. The details of the ration composition for both diet can be found in Salleh, et al. [6].

Biological samples

Liver biopsies (approximately 20 mg) were collected from each cow at the end of each feeding trial, RNA was extracted and sequenced. The details of the samples collection and processing were described in Salleh, et al. [6].

Blood samples were used for the DNA genotyping procedure. Ten milliliters blood samples were collected from the 19 cows using Ethylenediaminetetraacetic acid (EDTA) coated blood tubes. The blood samples were stored at -20 °C pending genomic DNA isolation and genotyping. The DNA was isolated and genotyped by Neogen GeneSeek® (Lincoln, NE, USA) using 777k BovineHD BeadChip (Illumina, Inc., San Diego, CA, USA).

Gene expressions data, genotype data and data pre-processing

Briefly, the RNA-seq data were pre-processed and processed to find candidate gene through differential expression analysis and weighted gene co-expression network analysis (WGCNA). RNA-seq analysis was performed as previously described in Salleh, et al. [6]. Briefly, the RNA raw reads were pre-processed using FastQC version 0.11.3 [19]. The reads were aligned to the Bovine reference genome release 82 using STAR aligner [20]. After the alignment, quality control of the mapped reads was done using Qualimap version 2.0 [21]. Then the HTSeq-count tool was used to compute the gene expression counts [22]. The DEGs analyses were done using DESeq2 package [23] and weighted gene co-expression analyses using WGCNA package [24]. Hub genes were selected from the top significant modules that have significant association with RFI trait and having more than 80% module membership. The RNA-seq data for the present study is available in

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92398.

Subsequently, the two gene expression datasets of Holsteins and Jerseys were preprocessed independently to filter low counts genes in each breed. Next, the two datasets were merged by keeping only genes present in both RNA-seq datasets.

We performed two separate analyses as replicates, one using the expression profile of the cows on C diet and another one using the expression profile of the cows on the HC diet. A summary of the eQTL mapping pipeline is presented in Figure 2. The eQTL mapping analysis was performed on 170 candidate genes for RFI (Supplementary Table 1, Supplementary Table 2, Supplementary Table 3 and Supplementary Table 4) that were identified in two previous studies based on the same RNA-seq data. The total list of candidate genes included 89 DEGs between cows with high and low RFI [6] and 81 hub genes in groups of co-expressed genes associated with RFI identified by using a weighted gene co-expression network analysis (WGCNA) (unpublished).

Among the 170 candidate genes, 160 survived after the filtering step in both datasets and were used in the rest of the analyses. The numbers of candidate genes that survived after filtration were the same in the two separate analyses, which were performed for cows fed low as compared to high concentrate diets. The log2 transformation of the gene count matrix was used in the eQTL mapping.

The genotype data was filtered by Hardy Weinberg

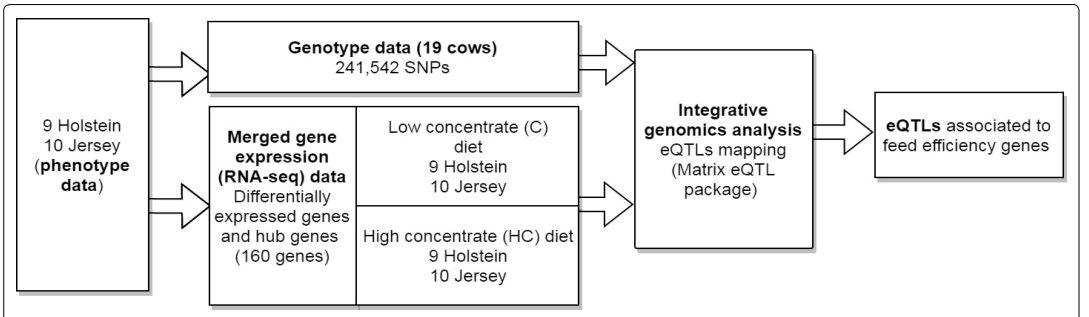


Figure 2: eQTL mapping pipeline used to find candidate genes for Residual Feed Intake (RFI) in dairy cattle. Among 170 candidate genes, only the 160 genes present in RNA-seq datasets for both breeds were used in the further analyses.

Equilibrium (HWE < 0.0001), Minor Allele Frequency (MAF < 0.15), and missing genotype rates (mind > 0.1). The genotype data were also pruned to remove SNPs in strong linkage disequilibrium [25]. The preprocessing was performed using PLINK 1.90 beta software [26]. A total of 536,420 SNPs was removed after the filtration procedure. The remaining 241,542 SNPs were used for the rest of the analysis.

Integrative genomics analysis (eQTL mapping)

The theoretical aspects of eQTL mapping and applications of findings in animal sciences are well described in the literature, including our previous studies [9-11]. The eQTL mapping was performed by fitting an analysis of variance (ANOVA) model to test both additive and dominant effects. The Matrix eQTL v2.1.1 [27] package in R software was used to identify the local and distant-eQTL associated to the RFI trait. We included the breed and the lactation number as covariates in the model.

The local-and distant-eQTLs analyses were performed separately. The analysis of local eQTL was performed on SNPs that were located at less than 1 Mb distance from the start or end position of the gene of interest, while distant-eQTLs analysis was performed on SNPs located at a distance of more than 1 Mb on the same chromosome and on SNPs in other chromosomes. The SNPs were mapped onto the *Bos taurus* genome UMD 3.1. The information about gene locations were retrieved from Ensembl database for *Bos taurus* v82. P-values were adjusted using the false discovery rate (FDR) procedure for multiple comparison corrections [28]. SNPs were considered significant with FDR lower than 0.05.

Comparison of the eQTL with the Animal Genome cattle QTLdb

The significant eQTLs identified in this study were further compared to the Animal Genome cattle QTLdb database [29]. From the cattle QTL database we filtered out long QTL regions and more than one flank markers. In total, 94,322 SNPs were used in the comparison. The SNPs information was obtained from 337 studies, 63 breeds, 366 traits of 6 trait types. The flanking regions of 500 kb around the eQTL identified in our studies were compared against the cattle QTLdb. The QTLs overlapping for at least one nucleotide were considered as a match.

Results and Discussion

The eQTL mapping analysis allows identification of SNPs associated with the expression level of specific genes. The hypothesis of this analysis is that the eQTL or eSNPs are in linkage with regulatory regions or region that encode for transcription factors responsible for the control of the expression of the targeted gene [10]. In the present study, we have analyzed candidate genes associated with the RFI trait in dairy cattle. Despite the small sample size, we identified several loci significantly related to the expression of the candidate genes. In addition, since the study focused on genes significantly associated with RFI, and the eQTL analysis was done on animals with widely different RFI, either very low or very high, our study had enough power to detect biologically meaningful expression variants.

Different strategies for obtaining significant eQTLs associated to RFI candidate genes by using C versus HC dataset

In the expression profile of cows fed the control diet, we identified 20 local-eQTLs SNPs or cis-eQTLs SNP (FDR < 0.05) associated with the expression of seven genes (*BDH2, CHRNE, ELOVL6, GIMAP4, FDXR, CXCL9* and *CD52*) (Table 1). However, there was no significant distant-eQTL (trans-eQTLs) associated with the candidate genes in the analysis performed among cows fed the control diet. On the other hand, among cows fed

Table 1: Top significant local-eQTLs targeting candidate genes for cows fed the control (C) diet.

rsID	Gene.name	Gene type	FDR	Position	Freq	Genotype	Gene position
rs133674837	*BDH2*	DE	7.35E-04	6:23051485	6/8/5	CC/CA/AA	6:23047057-23077431
rs109975461	*CHRNE*	DE	4.23E-02	19:26981374	8/6/5	AA/AG/GG	19:27118517-27123114
rs109947248	*CHRNE*	DE	4.23E-02	19:26994134	8/6/5	AA/AG/GG	19:27118517-27123114
rs109341116	*CHRNE*	DE	4.23E-02	19:27098154	8/6/5	AA/AG/GG	19:27118517-27123114
rs110896981	*CHRNE*	DE	4.23E-02	19:27192150	10/4/5	AA/AG/GG	19:27118517-27123114
rs43318602	*ELOVL6*	DE	4.23E-02	6:16656338	15/2/2	GG/GA/AA	6:16376642-16510240
rs43318545	*ELOVL6*	DE	4.23E-02	6:16678359	15/2/2	GG/GA/AA	6:16376642-16510240
rs43317462	*ELOVL6*	DE	4.23E-02	6:16731878	9/8/2	GG/GA/AA	6:16376642-16510240
rs110036492	*ELOVL6*	DE	4.23E-02	6:16738741	8/8/2	GG/GA/AA	6:16376642-16510240
rs43315610	*ELOVL6*	DE	4.23E-02	6:16755625	6/11/2	GG/GA/AA	6:16376642-16510240
rs43316358	*ELOVL6*	DE	4.23E-02	6:16761983	6/11/2	GG/GA/AA	6:16376642-16510240
rs43317449	*ELOVL6*	DE	4.23E-02	6:16725243	9/8/2	GG/GA/AA	6:16376642-16510240
rs109963253	*GIMAP4*	DE	3.02E-02	4:113638587	8/9/2	GG/GA/AA	4:113866800-113874303
rs134589272	*FDXR*	DE	4.23E-02	19: 56624163	10/8/1	GG/GA/AA	19: 57164031-57175524

rs ID = reference SNP cluster ID; Gene Name = name of the targeted genes; Gene type = type of candidate gene: Hub genes (HG) or differentially expressed genes (DE); FDR = False Discovery Rate of the association between SNP and gene expression, Position = genome position of the SNP, Freq = frequency of each genotypes (A1A1/A1A2/A2A2) in the cows analysed, Genotype = genotype at the SNP locus (A1A1/A1A2/A2A2), Gene position = genome position of the targeted gene.

Table 2: Top significant local-eQTLs targeting candidate genes for cows fed the high concentrate (HC) diet.

rsID	Gene.name	Gene type	FDR	Position	Freq	Genotype	Gene position
rs135948495	*UHRF1*	DE	6.72E-03	7:20322296	5/9/5	CC/CA/AA	7:20436670-20469912
rs134849198	*UHRF1*	DE	6.72E-03	7:20327318	5/9/5	CC/CA/AA	7:20436670- 20469912
rs137012774	*UHRF1*	DE	6.72E-03	7:20336175	5/9/5	CC/CA/AA	7:20436670- 20469912
rs133674837	*BDH2*	DE	4.81E-03	6:23051485	6/8/5	CC/CA/AA	6:23047057- 23077431
rs109739833	*HSD17B4*	DE	1.14E-02	7:35653128	7/10/2	AA/AG/GG	7:35662599- 35763653
rs110212970	*GIMAP4*	DE	7.88E-04	4:113608223	8/9/2	GG/GA/AA	4:113866800- 113874303
rs109963253	*GIMAP4*	DE	7.88E-04	4:113638587	8/9/2	GG/GA/AA	4:113866800- 113874303

rs ID = reference SNP cluster ID; Gene Name = name of the targeted genes; Gene type = type of candidate gene: Hub genes (HG) or differentially expressed genes (DE); FDR = False Discovery Rate of the association between SNP and gene expression, Position = genome position of the SNP, Freq = frequency of each genotypes (A1A1/A1A2/A2A2) in the cows analysed, Genotype = genotype at the SNP locus (A1A1/A1A2/A2A2), Gene position = genome position of the targeted gene.

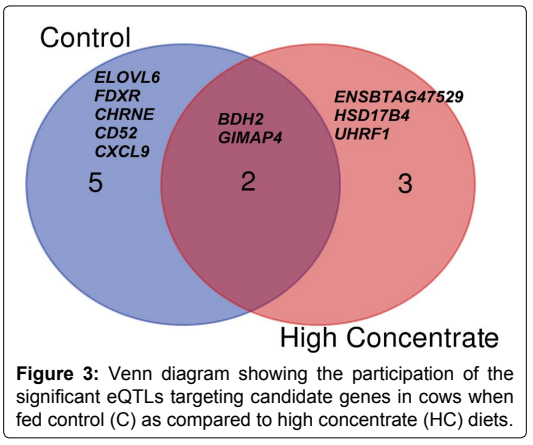


Figure 3: Venn diagram showing the participation of the significant eQTLs targeting candidate genes in cows when fed control (C) as compared to high concentrate (HC) diets.

the high concentrate diet, we identified 16 local eQTLs SNPs (FDR < 0.05) associated with the expression of five genes (*UHRF1, BDH2, HSD17B4, GIMAP4* and ENSBTAG00000047529) (Table 2) and 2891 distant-eQTLs associated with the expression of 45 genes. Among the local-eQTL, genes that were in common in both diet groups were the *BDH2* and *GIMAP4* genes. Figure 3 shows the significant eQTLs targeting candidate genes. A complete list of the significant distant-eQTLs, including the chromosomal position and annotation of the SNPs in the HC dataset analysis is presented in Supplementary Table 5.

In Holstein, cows fed with C diet dataset, we detected eQTLs associated to the three genes *BDH2, CHRNE* and *ELOVL6*, whereas for cows fed the HC diet dataset, the eQTLs associated to two other genes, *UHRF1* and *HSD17B4*. In Jersey cows, two DEGs (*GIMAP4* and *FDXR*) and two HG's (*CXCL9* and *CD52*) belonging to a group of co-expressed genes associated with RFI, when they were fed the C diet, whereas only the *GIMAP4* gene was detected as significant local-eQTL, when they were fed the HC diet. However, the HG's (*CXCL9* and *CD52*) were not further analyzed, since only one animal with rare allele at these loci were present in the dataset. Supplementary Figure 1 and Supplementary Figure 2 present boxplots of genotypes and their correlation with gene expressions for the top seven significant local-eQTLs corresponding to seven candidate genes (*BDH2, CHRNE, ELOVL6, GIMAP4, FDXR, UHRF1* and *HSD17B4*) for the RFI trait in the present study.

Previous studies showed that identification of eQTLs and genomic regions would give additional information towards the identification of causal variants [30]. Hence, the eQTLs that were identified as associated to the RFI trait in the present study would provide additional information for the development of biomarkers.

The first top eQTL with a significant relationship between the gene expression and genotype is rs133674837, which is associated to the *BDH2* gene (Supplementary Figure 1a), and as mentioned the association was found to be significant in the two separate analyses for cows when fed the C diet as well as when fed the HC diet. The expression of the *BDH2* gene was previously identified to be upregulated in high FE cows [6]. All low RFI (high FE) Holstein cows (n = 5) had homozygous (AA) genotype at this locus, while 80% of the high RFI (low FE) Holstein cow had heterozygous genotype (CA). *BDH2* encodes for the enzyme 3-Hydroxybutyrate Dehydrogenase 2, which is responsible for degradation of 3-hydroxybutyrate-a ketone body derived partly from rumen fermentation and partly from incomplete oxidation of fatty acids in the liver [31,32]. The *BDH2* gene in the liver has been observed to be downregulated in animals, when ketogenesis occurred (mice and pigs) [31,33]. This happens, for example, during feed restriction or fasting of animals, and mRNA expression of *BDH2* gene has been shown to be lower in such animals compared to normal feeding animals [33]. In the present study, the hepatic *BDH2* gene expression was downregulated in high RFI (low FE) animals. The positive association between a homozygous (AA) genotype and upregulation of the *BDH2* gene in low RFI (high FE) Holstein cows shows that this locus might influence the RFI trait. However, in Jersey cows, 80% (n = 4) of low RFI (high FE) cows were homozygous (CC) at this allele. Hence, specifically for Holstein cows, a homozygous (AA) genotype is expected to favor low RFI and hence high FE.

GIMAP4 gene is another gene that has been detected as significantly associated with the eQTLs listed in Table 1 and Table 2 in both analysis (i.e. when cows were fed

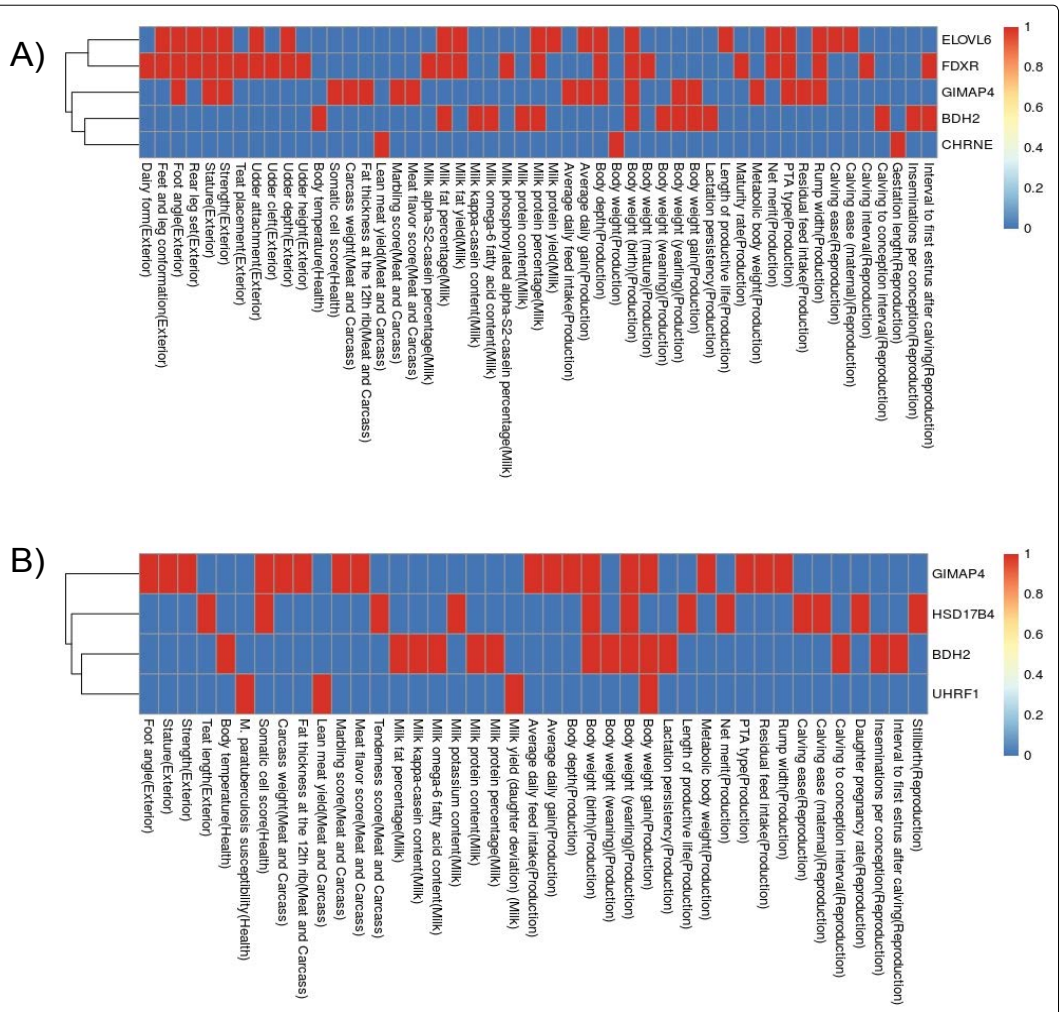


Figure 4: Heatmap showing SNPs corresponding to seven genes overlapping with previous QTLs for major traits in cattle QTL database (a) Local-eQTLs associated with genes in cows on the control (C) diet analysis; (b) Local-eQTLs associated with genes in cows on the hogh concentrate (HC) diet. Red: with at least one hit; Blue: no hit.

the C and HC diets). The top significant eQTLs targeting *GIMAP4* was rs109963253. All the five low RFI (high FE) Jersey cows were heterozygous (GA) at this SNP locus, while 60% (n = 3) of the high RFI (low FE) cows were homozygous (GG) (Supplementary Figure 1b). GIMAP4 encodes for a GTPase binding protein responsible for regulating lymphocyte apoptosis (http://www.genecards.org/cgi-bin/carddisp.pl?gene=GIMAP4). Hence in humans, the *GIMAP4* gene has been shown to be involved in immunological responses [34]. We already in our previous paper [6] stated the importance of the *GIMAP* genes for the FE trait, since the *GIMAP4* gene was significantly higher expressed in high feed efficient than low feed efficient Jersey cows. The present study is thus in line with conclusions from previous studies, where function of immunological responses has been associated to productivity and FE in farm animals [35-38]. A possible explanation is that animals with poorer immune function are more prone to develop infectious diseases like e.g. mastitis, and this can reduce milk production, induce fever-associated increases in metabolism, and hence increased energy expenditures per kg of produced milk, which subsequently reduces FE [39].

rs109975461, which is associated with the *CHRNE* gene, was also a significant eQTL. At this locus, all high RFI (low FE) Holstein cows had a homozygous (GG) genotype, whereas 80% of the low RFI (high FE) cows (n = 4) had a heterozygous (AG) genotype (Supplementary Figure 1c). In other words, high feed efficient cows that had a high expression of the *CHRNE* gene also had the heterozygote genotype. However, in the Jersey group, there was no association to be seen for this *CHRNE* gene. The *CHRNE* gene encodes for the acetylcholine recep-

tors in mature mammalian neuromuscular junctions. In general, this gene was never discussed before in relation with FE traits. Acetylcholine has been reported to influence hepatocyte glucose metabolism in rodents via actions on muscarinic receptors [40], but whether this is also the case in ruminants is not clear. Perhaps, more importantly, acetylcholine plays a critical role in the complex regulation of hypothalamic neuronal activity that influences feed intake [41], and in dairy cows, feed intake is a major factor limiting milk production in high-yielding dairy cows in early lactation [42].

Another interesting candidate identified in the analyses of Holstein cows on the C diet was the *ELOVL6* gene. In our study, the top SNP targeting *ELOVL6* gene was rs43315610. The *ELOVL6* gene has previously been discovered as an important gene that influences FE in beef cattle and pigs [43,44]. *ELOVL6*, which is also known as elongation of very long chain fatty acids protein 6, is part of the pathway of de novo fatty acid synthesis [45]. The lower expression of this gene in low RFI Holstein cows might be associated with low rates of de novo synthesis of fatty acids, as it has previously been described in pigs [44,46], and de novo synthesized fatty acids constitute up to 50% of fatty acids in milk on a molecular weight basis [47]. This gene has also been associated to long chain fatty acid synthesis in beef cattle [48]. Previously, the expression of *ELOVL6* was found differentially expressed in liver, adipose tissue and muscle [48]. In another study on QTL mapping for RFI in Holstein calves, it was found that another gene involved in fatty acid metabolism, *FABP4* gene were significantly associated with the top SNPs significantly associated RFI across three stages of age [49]. This gene is encoded for fatty acid binding protein which suggests that fatty acid synthesis and metabolism may be important parts of the RFI trait. In Jersey cattle, we did not observe any relation between RFI genotype and the *ELOVL6* gene expression. We found that 80% (n = 4) of the low RFI (high FE) Holstein cows had a heterozygous (GA) genotype, while 20% (n = 1) were homozygous (GG) (Supplementary Figure 1d). Therefore, a heterozygous genotype is expected to favor high FE.

When cows were fed the C diet, rs134589272 was identified as an eQTL, which corresponded to the *FDXR* gene in Jersey cows. All (n = 5) low RFI (high FE) Jersey cows were heterozygous (GA) and had high expression of this gene, while 80% (n = 4) of the high RFI (low FE) Jersey cows were homozygous (GG) corresponding to a lower expression of the *FDXR* gene (Supplementary Figure 1e). For Holstein cows, RFI was not related neither to this eQTL nor to the genotype for the *FDXR* gene. Functions of the *FDXR* gene are related to cholesterol metabolism [6], which is an important feature of e.g. membrane synthesis, which is important for formation of the milk fat globule membrane covering secreted milk fat.

In addition, another four eQTLs associated to the gene *HSD17B4* and *UHRF1* expression were found as significant in the analysis for cows fed the HC diet (Supplementary Figure 2a and Supplementary Figure 2b). Interestingly, these two genes were also previously found associated with the FE trait. The *HSD17B4* gene encodes for a major enzyme involved in peroxisomal β-oxidation, and it was found to be upregulated in abdominal fat of low growth chicken [50], and this appears to be in line with the present study, where the *HSD17B4* gene expression was upregulated in the high RFI Holstein cows. *UHRF1* encodes for Ubiquitin like With PHD and Ring Finger Domains 1 (http://www.genecards.org/cgi-bin/carddisp.pl?gene=UHRF1), which is an essential regulator of DNA methylation. Several studies have identified that Ubiquitin family genes were significantly associated with RFI traits in *Bos taurus* [51,52]. In the present study *UHRF1* gene was found significantly downregulated in high RFI Holstein cows.

Overlapping genomic regions for FE trait in the QTL database

In order to gain more information regarding the eQTLs that we discovered in the present study, we compared the results of the SNPs locations with the previously reported QTLs and variants from GWAS study from the Animal genome cattle QTL database. We identified several overlaps of our eQTLs with QTLs from previous studies. The QTLs overlapping with our eQTLs were associated with a different type of traits (Figure 4a and Figure 4b).

The eQTLs which associated to the expression of *ELOVL6* and *FDXR* genes are the most overlapped with many traits. Only the *GIMAP4* gene was previously associated to production traits, such as RFI, rump width, metabolic body weight, body weight gain, body weight (yearling), body weight, body depth, average daily gain as well as average daily feed intake [53,54]. However, the same region contains QTLs for other traits, such as reproduction, milk, meat and carcass, health and exterior association traits [55,56].

The fact that all these associations with different type of traits were found within this 1Mb region, shows that this must be a significant region with control points for several targeting genes. The eQTLs identified are close to the QTL for production traits and for FE traits. At the same time, this confirms that the candidate genes which associated to FE trait in our findings were also closely associated to several production traits. However, the association with other important traits can be a sign of double association between reproduction and production traits, which were well discussed elsewhere [57]. Thus, the uses of genomic region information need to be tested and validated in a different and a larger population before further usage in any genomic selection procedures can be implemented.

Conclusion

To bridge the gap between genotype and phenotype, we attempted in this study to identify DEGs and HG's among previously identified candidate genes for the FE trait. The identified local-eQTLs provide additional evidence of the involvement of some of previously identified candidate genes in RFI determination, and our study provides new information on possible regulatory and causative genetic variants that can be used in genomics-based selection for FE in dairy cows. We identified eQTL associated to the expression of seven genes (*BDH2, CHRNE, ELOVL6, GIMAP4, UHRF1, HSD17B4 and FDXR*) that appear to be involved in metabolic pathways related to RFI and hence feed efficiency. The eQTLs overlapped with QTLs previously associated with FE trait (e.g. dry matter intake, longevity, body weight gain and net merit). Interestingly, Holstein and Jersey cows appear to rely on different strategies to achieve low RFI, and this was associated to cholesterol and lipid metabolism related pathways in Holstein cows, but to immune and inflammatory related functions in Jersey cows. Thus, our findings suggest that the identified eQTLs can be used as potential biomarkers for feed efficiency and used to predict feed efficiency level. The genomic region around the identified SNP markers could be included in genomics/genetic-based selection in Holstein and Jersey. However, before applying this new knowledge in genetic testing or in commercial applications, the results must be validated in a larger population, and it must be further analyzed if pleiotropic effects of eQTLs also include adverse disease traits.

Acknowledgements

The authors would like to thank the collaborator scientists, specifically Peter Lund, Johanna Höglund and Dana Olijhoek, for contributions to the present study. This project was funded by Milk Levy Foundation, Skejby Denmark. The PhD student (Suraya Mohamad Salleh) would also like to acknowledge the Ministry of Higher Education of Malaysia and Universiti Putra Malaysia for the PhD stipend as well as University of Copenhagen for the tuition fee waiver.

References

1. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet 10: 57-63.
2. Weber KL, Welly BT, Van Eenennaam AL, Young AE, Porto-Neto LR, et al. (2016) Identification of gene networks for residual feed intake in angus cattle using genomic prediction and RNA-seq. PLoS One 11: e0152274.
3. Alexandre PA, Kogelman LJ, Santana MH, Passarelli D, Pulz LH, et al. (2015) Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. BMC Genomics 16: 1.
4. Paradis F, Yue S, Grant J, Stothard P, Basarab J, et al. (2015) Transcriptomic analysis by RNA sequencing reveals that hepatic interferon-induced genes may be associated with feed efficiency in beef heifers. J Anim Sci 93: 3331-3341.
5. Koch RM, Swiger LA, Chambers D, Gregory KE (1963) Efficiency of Feed Use in Beef Cattle. Journal of Animal Science 22.
6. Salleh M, Mazzoni G, Höglund J, Olijhoek D, Lund P, et al. (2017) RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high-and low-residual feed intake in Nordic dairy cattle. BMC Genomics 18: 258.
7. Le Mignon G, Désert C, Pitel F, Leroux S, Demeure O, et al. (2009) Using transcriptome profiling to characterize QTL regions on chicken chromosome 5. BMC Genomics 10: 575.
8. Civelek M, Lusis AJ (2014) Systems genetics approaches to understand complex traits. Nat Rev Genet 15: 34.
9. Kadarmideen HN, Reverter A (2007) Combined genetic, genomic and transcriptomic methods in the analysis of animal traits. CABI review: Perspectives in Agriculture, Veterinary science, Nutrition and Natural resources 2: 16.
10. Kadarmideen HN, von Rohr P, Janss LL (2006) From genetical genomics to systems genetics: Potential applications in quantitative genomics and animal breeding. Mamm Genome 17: 548-564.
11. Suravajhala P, Kogelman LJ, Kadarmideen HN (2016) Multi-omic data integration and analysis using systems genomics approaches: Methods and applications in animal production, health and welfare. Genet Sel Evol 48: 38.
12. Kadarmideen H (2008) Genetical systems biology in livestock: Application to gonadotrophin releasing hormone and reproduction. IET Syst Biol 2: 423-441.
13. Mazzoni G, Pedersen H, De Oliveira Junior G, Alexandre P, Razza E, et al. (2017) Application of integrative genomics and systems biology to conventional and in vitro reproductive traits in cattle.14.
14. Yuan J, Wang K, Yi G, Ma M, Dou T, et al. (2015) Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. Genetics Selection Evolution 47: 82.
15. Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, et al. (2017) Genome-wide association study for feed efficiency and growth traits in US beef cattle. BMC genomics 18: 386.
16. Rolf M, Taylor J, Schnabel R, McKay S, McClure M, et al. (2012) Genome-wide association analysis for feed efficiency in Angus cattle. Anim Genet 43: 367-374.
17. De Koning D, Cabrera C, Haley C (2007) Genetical genomics: Combining gene expression with marker genotypes in poultry. Poult Sci 86: 1501-1509.
18. Li B, Fikse W, Lassen J, Lidauer M, Løvendahl P, et al. (2016) Genetic parameters for dry matter intake in primiparous Holstein, Nordic Red, and Jersey cows in the first half of lactation. J Dairy Sci 99: 7232-7239.
19. Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. In.
20. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15-21.
21. Okonechnikov K, Conesa A, García-Alcalde F (2016) Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32: 292-294.
22. Anders S, Pyl PT, Huber W (2014) HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166-169.

23. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550.

24. Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9: 559.

25. Dudbridge F, Newcombe PJ (2015) Accuracy of gene scores when pruning markers by linkage disequilibrium. Hum Hered 80: 178-186.

26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, et al. (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4: 7.

27. Shabalin AA (2012) Matrix eQTL: Ultra-fast eQTL analysis via large matrix operations. Bioinformatics 28: 1353-1358.

28. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B (Methodological) 289-300.

29. Hu ZL, Park CA, Reecy JM (2015) Developmental progress and current status of the Animal QTLdb. Nucleic Acids Res 44: 827-833.

30. Kogelman LJ, Cirera S, Zhernakova DV, Fredholm M, Franke L, et al. (2014) Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. BMC Med Genomics 7.

31. Geisler C, Hepler C, Higgins M, Renquist B (2016) Hepatic adaptations to maintain metabolic homeostasis in response to fasting and refeeding in mice. Nutr Metab 13: 62.

32. Khanal P, Axel AM, Kongsted AH, Husted SV, Johnsen L, et al. (2015) Late gestation under-and overnutrition have differential impacts when combined with a post-natal obesogenic diet on glucose-lactate-insulin adaptations during metabolic challenges in adolescent sheep. Acta Physiol 213: 519-536.

33. Grala TM, Kay JK, Phyn CV, Bionaz M, Walker CG, et al. (2013) Reducing milking frequency during nutrient restriction has no effect on the hepatic transcriptome of lactating dairy cattle. Physiol Genomics 2013.

34. Heinonen M (2015) Players of Th-cell differentiation and immune dysregulation–focus on GIMAP family genes.

35. Banos G, Wall E, Coffey MP, Bagnall A, Gillespie S, et al. (2013) Identification of immune traits correlated with dairy cow health, reproduction and productivity. PLoS One 8: e65766.

36. Patience JF, Rossoni-Serão MC, Gutiérrez NA (2015) A review of feed efficiency in swine: Biology and application. J Anim Sci Biotechnol 6: 33.

37. Gondret F, Vincent A, Houée-Bigot M, Siegel A, Lagarrigue S, et al. (2017) A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs. BMC Genomics 18: 244.

38. Mpetile Z (2014) Effect of divergent selection for residual feed intake on immune system of Yorkshire pigs.

39. Johnson R (2012) Fueling the immune response: What's the cost? In: Feed efficiency in swine. Springer 211-223.

40. Vatamaniuk M, Horyn O, Vatamaniuk O, Doliba N (2003) Acetylcholine affects rat liver metabolism via type 3 muscarinic receptors in hepatocytes. Life Sci 72: 1871-1882.

41. Picciotto MR, Higley MJ, Mineur YS (2012) Acetylcholine as a neuromodulator: Cholinergic signaling shapes nervous system function and behavior. Neuron 76: 116-129.

42. Johansen M, Lund P, Weisbjerg MR (2018) Feed intake and milk production in dairy cows fed different grass and legume species: A meta-analysis. Animal 12: 66-75.

43. Lemos MV, Chiaia HLJ, Berton MP, Feitosa FL, Aboujaoud C, et al. (2016) Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. BMC Genomics 17: 213.

44. Reyer H, Oster M, Magowan E, Dannenberger D, Ponsuksili S, et al. (2017) Strategies towards improved feed efficiency in pigs comprise molecular shifts in hepatic lipid and carbohydrate metabolism. Int J Mol Sci 18: 1674.

45. Moon YA, Ochoa CR, Mitsche MA, Hammer RE, Horton JD (2014) Deletion of ELOVL6 blocks the synthesis of oleic acid but does not prevent the development of fatty liver or insulin resistance. J Lipid Res 55: 2597-2605.

46. Corominas J, Marchesi JA, Puig-Oliveras A, Revilla M, Estellé J, et al. (2015) Epigenetic regulation of the ELOVL6 gene is associated with a major QTL effect on fatty acid composition in pigs. Genet Sel Evol 47: 20.

47. Breckenridge W, Kuksis A (1967) Molecular weight distributions of milk fat triglycerides from seven species. J Lipid Res 8: 473-478.

48. Chen S, He H, Liu X (2017) Tissue expression profiles and transcriptional regulation of elongase of very long chain fatty acid 6 in bovine mammary epithelial cells. PLoS One 12: e0175777.

49. Cohen-Zinder M, Asher A, Lipkin E, Feingersch R, Agmon R, et al. (2016) FABP4 is a leading candidate gene associated with residual feed intake in growing Holstein calves. Physiol Genomics 48: 367-376.

50. Resnyk CW, Carré W, Wang X, Porter TE, Simon J, et al. (2017) Transcriptional analysis of abdominal fat in chickens divergently selected on bodyweight at two ages reveals novel mechanisms controlling adiposity: Validating visceral adipose tissue as a dynamic endocrine and metabolic organ. BMC Genomics 18: 626.

51. Karisa B, Moore S, Plastow G (2014) Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle. Anim Sci J 85: 374-387.

52. de Oliveira PS, Cesar AS, do Nascimento ML, Chaves AS, Tizioto PC, et al. (2014) Identification of genomic regions associated with feed efficiency in Nelore cattle. BMC Genet 15: 100.

53. Lu D, Miller S, Sargolzaei M, Kelly M, Vander Voort G, et al. (2013) Genome-wide association analyses for growth and feed efficiency traits in beef cattle. J Anim Sci 91: 3612-3633.

54. Snelling W, Allan M, Keele J, Kuehn L, Mcdaneld T, et al. (2010) Genome-wide association study of growth in crossbred beef cattle 1 2. J Anim Sci 88: 837-848.

55. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, et al. (2011) Genome-wide association analysis of thirty-one production, health, reproduction and body conformation traits in contemporary US Holstein cows. BMC Genomics 12: 408.

56. Lindholm-Perry AK, Kuehn LA, Rempel LA, Smith TP, Cushman RA, et al. (2012) Evaluation of bovine chemerin (rarres2) gene variation on beef cattle production traits1. Front Genet 3: 39.

57. Dobson H, Smith R, Royal M, Knight C, Sheldon I (2007) The high-producing dairy cow and its reproductive performance. Reprod Domest Anim 42: 17-23.

Supplementary Table 1: List of differentially expressed genes for Holstein.

	Ensembl.Gene.ID	Associated.Gene.Name	baseMean	log2FoldChange	padj
1	ENSBTAG00000000170	Uncharacterized protein	226.9268	-0.42496	0.013405
2	ENSBTAG00000000654	*ARMC4*	58.66466	-0.59056	0.000358
3	ENSBTAG00000001009	*HCLS1*	440.5817	0.322596	0.028031
4	ENSBTAG00000001154	*DGAT2*	511.454	-0.36983	0.0292
5	ENSBTAG00000001204	*KIAA1462*	225.2099	-0.41583	0.017639
6	ENSBTAG00000002224	*UHRF1*	77.58827	-0.49601	0.000209
7	ENSBTAG00000002526	*BDH2*	1382.085	-0.57884	3.67E-16
8	ENSBTAG00000002705	*REC8*	304.6793	-0.36531	0.00171
9	ENSBTAG00000003696	*CCDC64*	45.09122	0.445439	0.026992
10	ENSBTAG00000003718	*HACL1*	6329.552	0.315377	0.038742
11	ENSBTAG00000004076	*OXER1*	223.6781	-0.43485	0.010458
12	ENSBTAG00000004558	*C15orf48*	89.05707	0.504892	0.003366
13	ENSBTAG00000004908	*CHRNE*	246.4988	-0.74877	6.38E-08
14	ENSBTAG00000005287	*CYP7A1*	4126.209	0.456229	0.013057
15	ENSBTAG00000005629	*AIM1L*	913.5876	-0.29846	0.003366
16	ENSBTAG00000006452	*CD3D*	77.7469	0.40702	0.036637
17	ENSBTAG00000006599	*INHBE*	605.9473	-0.4237	0.044088
18	ENSBTAG00000006675	*PCSK6*	3039.353	-0.1866	0.028031
19	ENSBTAG00000006934	*CYP11A1*	649.8806	0.48638	0.004844
20	ENSBTAG00000006978	*HSD17B4*	13797.37	0.301172	0.026992
21	ENSBTAG00000006999	*RYR1*	148.3439	0.516872	0.001988
22	ENSBTAG00000007554	*IFI6*	136.8134	0.375726	0.049041
23	ENSBTAG00000007828	*SLA*	118.2561	0.322695	0.038742
24	ENSBTAG00000007895	*SLC20A1*	880.0804	-0.56325	1.36E-05
25	ENSBTAG00000008160	*MBOAT2*	440.5952	0.344004	0.043864
26	ENSBTAG00000008424	*ABR*	459.6035	0.329325	0.031825
27	ENSBTAG00000008913	*TMEM98*	333.1394	-0.51804	0.001004
28	ENSBTAG00000009085	*SLC35A5*	1691.358	0.277516	0.006652
29	ENSBTAG00000009137	*NKG7*	215.4444	0.380987	0.028031
30	ENSBTAG00000009263	*MFSD1*	2661.739	0.240088	0.013057
31	ENSBTAG00000010463	Uncharacterized protein	394.1386	0.383919	0.002458
32	ENSBTAG00000010564	*ELOVL6*	994.7373	0.43905	0.026992
33	ENSBTAG00000011771	*FICD*	107.9076	-0.35997	0.037153
34	ENSBTAG00000011832	*ALDH18A1*	404.9092	0.313185	0.025962
35	ENSBTAG00000012007	*SOCS2*	835.3211	0.422255	0.043864
36	ENSBTAG00000012995	*CCDC109B*	52.93337	0.424138	0.037153
37	ENSBTAG00000013596	*NR1H4*	1215.489	0.241071	0.011983
38	ENSBTAG00000014064	*FGFR2*	1554.53	-0.40189	0.001141
39	ENSBTAG00000014791	*CTH*	224.7532	-0.5345	1.74E-06
40	ENSBTAG00000015313	*CEACAM19*	51.60282	-0.94436	1.81E-14
41	ENSBTAG00000015419	*ARHGEF37*	204.9786	0.471594	0.001988
42	ENSBTAG00000016542	*LAMB3*	1783.587	0.424674	0.024742
43	ENSBTAG00000017567	*ACACA*	844.7716	0.404374	0.005072
44	ENSBTAG00000018116	*MTFP1*	88.16198	-0.38701	0.025962
45	ENSBTAG00000018548	*INTS7*	6522.273	0.238158	0.001988
46	ENSBTAG00000018604	*SEMA4G*	4847.004	-0.16063	0.049041
47	ENSBTAG00000018723	*SLC25A34*	96.51437	-0.44163	0.0292
48	ENSBTAG00000019585	*MYOM1*	962.386	0.453079	0.025089

49	ENSBTAG00000020116	*JSP.1*	2041.873	0.323612	0.016775
50	ENSBTAG00000020371	*ACOT8*	312.3978	0.417754	0.013405
51	ENSBTAG00000020375	Uncharacterized protein	5129.183	0.421767	0.031078
52	ENSBTAG00000020499	Uncharacterized protein	68.21442	0.571805	0.000358
53	ENSBTAG00000020755	*SELP*	478.8212	-0.39216	0.014977
54	ENSBTAG00000021746	*ANXA5*	333.8561	-0.38421	0.038742
55	ENSBTAG00000023851	*FAM102A*	229.0066	-0.50952	0.001422
56	ENSBTAG00000023929	*FOSL2*	189.6217	0.424473	0.028031
57	ENSBTAG00000024044	*CDKL4*	82.11652	0.51475	0.001988
58	ENSBTAG00000025258	Uncharacterized protein	102.6902	0.544344	0.00171
59	ENSBTAG00000025898	*TBC1D8*	442.302	0.271973	0.047591
60	ENSBTAG00000026779	*LYZ*	516.8438	0.643794	6.78E-06
61	ENSBTAG00000030966	*TAF6*	419.5936	-0.25649	0.010458
62	ENSBTAG00000035998	*CKB*	332.0427	0.385047	0.049041
63	ENSBTAG00000037913	Uncharacterized protein	436.5267	0.212286	0.042867
64	ENSBTAG00000037917	*SLC17A1*	2786.405	0.438054	0.017094
65	ENSBTAG00000038496	*CR2*	1355.757	-0.54159	3.72E-06
66	ENSBTAG00000038962	*SLC6A11*	2637.353	-0.37017	0.010028
67	ENSBTAG00000039731	*RND3*	1761.21	-0.25436	0.028031
68	ENSBTAG00000046076	Uncharacterized protein	124.7543	-0.41978	0.04847
69	ENSBTAG00000046730	Uncharacterized protein	139.8274	0.365363	0.049041
70	ENSBTAG00000047529	Uncharacterized protein	110.8376	-0.53501	0.001896

+v e log2 fold change = upregulated in low feed efficiency group; - ve log2 fold change = downregulated in low feed efficiency group.

Supplementary Table 2: List of hub genes for Holstein.

	Ensembl gene ID	Gene name	Module membership	Gene significance
1	ENSBTAG00000000197	*TRMT10A*	0.801	0.576
2	ENSBTAG00000001774	*SPRY2*	-0.814	-0.52
3	ENSBTAG00000001950	*RDH11*	0.852	0.441
4	ENSBTAG00000002412	*CYB5B*	0.907	0.633
5	ENSBTAG00000002435	*PTPRE*	0.852	0.767
6	ENSBTAG00000002714	*GNAI1*	0.901	0.557
7	ENSBTAG00000002827	*ACAT2*	0.946	0.691
8	ENSBTAG00000002966	*DNAJC13*	0.813	0.71
9	ENSBTAG00000003068	*MSMO1*	0.852	0.579
10	ENSBTAG00000003305	*NCF1*	0.802	0.642
11	ENSBTAG00000003696	*CCDC64*	0.837	0.679
12	ENSBTAG00000003718	*HACL1*	0.854	0.705
13	ENSBTAG00000003948		0.919	0.559
14	ENSBTAG00000004075	*IDI1*	0.87	0.607
15	ENSBTAG00000004688	*DHCR24*	0.859	0.555
16	ENSBTAG00000005183	*MVK*	0.906	0.497
17	ENSBTAG00000005498	*SQLE*	0.816	0.442
18	ENSBTAG00000005650	*SKAP2*	0.826	0.589
19	ENSBTAG00000005976	*HSD17B7*	0.809	0.55
20	ENSBTAG00000006999	*RYR1*	0.929	0.763
21	ENSBTAG00000007014	*CEP63*	0.823	0.623
22	ENSBTAG00000007079	*LCP1*	0.806	0.583
23	ENSBTAG00000007840	*HMGCR*	0.888	0.522
24	ENSBTAG00000007844	*CETN2*	0.836	0.335

25	ENSBTAG00000008160	*MBOAT2*	0.865	0.534
26	ENSBTAG00000008329	*CYTIP*	0.823	0.477
27	ENSBTAG00000010347	*EZR*	0.85	0.506
28	ENSBTAG00000011146	*RAB8B*	0.884	0.473
29	ENSBTAG00000011839	*HMGCS1*	0.871	0.507
30	ENSBTAG00000012059	*MVD*	0.831	0.364
31	ENSBTAG00000012170	*UBL3*	0.813	0.729
32	ENSBTAG00000012432	*FDFT1*	0.821	0.529
33	ENSBTAG00000012695	*LCK*	0.837	0.534
34	ENSBTAG00000013284		0.886	0.736
35	ENSBTAG00000013303	*ACSS2*	0.866	0.571
36	ENSBTAG00000013749	*RHOQ*	0.868	0.525
37	ENSBTAG00000014517	*KLB*	0.857	0.64
38	ENSBTAG00000015327	*SPTAN1*	0.899	0.637
39	ENSBTAG00000015980	*FASN*	0.859	0.49
40	ENSBTAG00000016445	*YME1L1*	0.807	0.717
41	ENSBTAG00000016465	*DHCR7*	0.903	0.521
42	ENSBTAG00000016709	*NT5C3A*	0.824	0.615
43	ENSBTAG00000016721	*ZNF791*	0.824	0.559
44	ENSBTAG00000016740	*ACLY*	0.918	0.52
45	ENSBTAG00000018936	*LSS*	0.839	0.58
46	ENSBTAG00000018959	*RAB11A*	0.828	0.67
47	ENSBTAG00000020984	*RAPGEF4*	0.856	0.775
48	ENSBTAG00000021842		0.804	0.492
49	ENSBTAG00000030951		0.844	0.508
50	ENSBTAG00000036260	*LPXN*	0.801	0.391
51	ENSBTAG00000037413	*TMEM164*	0.81	0.468
52	ENSBTAG00000047970		0.835	0.558

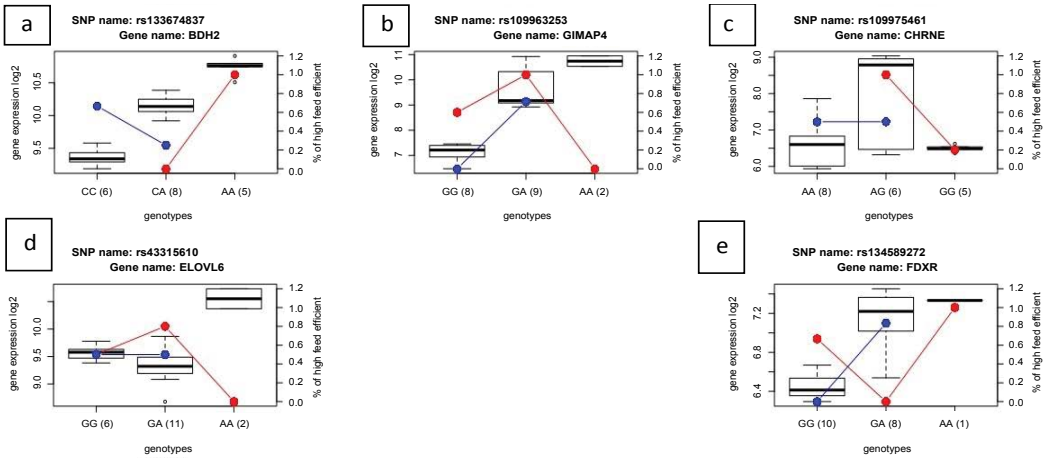
Supplementary Table 3: List of differentially expressed genes for Jersey.

	Ensembl.Gene.ID	Associated.Gene.Name	baseMean	log2FoldChange	padj
1	ENSBTAG00000006525	*FDXR*	125.97	-0.64501	6.21E-13
2	ENSBTAG00000008066	*PKDREJ*	76.74951	0.561566	1.15E-05
3	ENSBTAG00000013689	*MCTP2*	148.2616	0.528627	1.07E-05
4	ENSBTAG00000027727	Uncharacterized protein	284.1996	0.479634	0.000373
5	ENSBTAG00000038487	*ZNF613*	155.3981	-0.39026	0.026308
6	ENSBTAG00000046257	*GIMAP4*	650.1794	-0.38726	0.0024
7	ENSBTAG00000005182	*BOLA-A*	434.9538	-0.38721	0.001332
8	ENSBTAG00000014402	*GIMAP8*	713.0488	-0.38164	0.008566
9	ENSBTAG00000045727	Uncharacterized protein	921.1041	0.380952	0.032465
10	ENSBTAG00000019026	*EXTL2*	34.58975	0.377761	0.03724
11	ENSBTAG00000037440	*ZNF197*	281.1009	0.357972	0.016391
12	ENSBTAG00000021751	*RASEF*	36.02273	-0.3504	0.010571
13	ENSBTAG00000027205	*PGBD5*	30.06234	-0.34041	0.026308
14	ENSBTAG00000031737	*TMEM102*	26.52997	0.338561	0.03724
15	ENSBTAG00000009087	*GNG10*	1516.44	-0.32465	0.026308
16	ENSBTAG00000040323	Uncharacterized protein	1003.599	-0.32071	0.026308
17	ENSBTAG00000014161	*ARMC10*	258.84	-0.2983	0.026308
18	ENSBTAG00000013106	*C19orf81*	26.34468	0.295346	0.026308
19	ENSBTAG00000047379	*CYP3A4*	2422.366	0.286561	0.043386

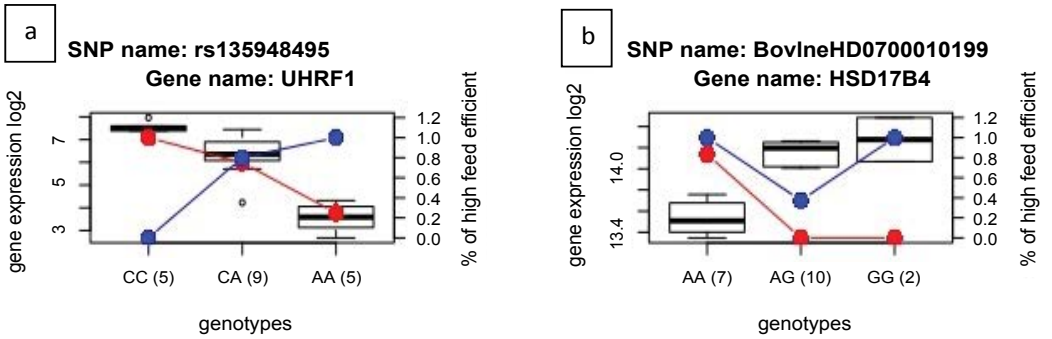
+ ve log2 fold change = upregulated in low feed efficiency group; - ve log2 fold change = downregulated in low feed efficiency group.

Supplementary Table 4: List of hub genes for Holstein.

	Ensembl gene ID	Gene name	Module membership	Gene significance
1	ENSBTAG00000000431	*TRDC*	0.858	-0.411
2	ENSBTAG00000000432	*TRAC*	0.86	-0.526
3	ENSBTAG00000000715		0.889	-0.487
4	ENSBTAG00000001198		0.81	-0.555
5	ENSBTAG00000002669	*RASSF4*	0.802	-0.722
6	ENSBTAG00000003037		0.829	-0.485
7	ENSBTAG00000004894		0.907	-0.497
8	ENSBTAG00000004917	*KLRK1*	0.826	-0.437
9	ENSBTAG00000005628		0.818	-0.49
10	ENSBTAG00000005892	*ZAP70*	0.864	-0.609
11	ENSBTAG00000006452	*CD3D*	0.9	-0.494
12	ENSBTAG00000006552	*LAMP3*	0.827	-0.501
13	ENSBTAG00000007191	*CCL5*	0.909	-0.48
14	ENSBTAG00000008401	*PFKFB3*	0.808	-0.547
15	ENSBTAG00000009381	*LCP2*	0.857	-0.654
16	ENSBTAG00000012695	*LCK*	0.852	-0.51
17	ENSBTAG00000013730	*CD5*	0.857	-0.403
18	ENSBTAG00000014725	*CD27*	0.822	-0.474
19	ENSBTAG00000015708	*CXCR6*	0.879	-0.469
20	ENSBTAG00000015710	*CD3E*	0.875	-0.537
21	ENSBTAG00000017256	*CD2*	0.914	-0.474
22	ENSBTAG00000019403	*MALSU1*	0.8	-0.536
23	ENSBTAG00000020904	*JAK3*	0.857	-0.439
24	ENSBTAG00000027246	*UBD*	0.888	-0.621
25	ENSBTAG00000030426		0.889	-0.379
26	ENSBTAG00000037510		0.853	-0.433
27	ENSBTAG00000038639	*CXCL9*	0.906	-0.425
28	ENSBTAG00000039588		0.815	-0.535
29	ENSBTAG00000047988		0.842	-0.365



Supplementary Figure 1: a-e) The boxplots show the five significant eQTLs with the associated genes for control diet group analysis. X-axis: genotypes; y-axis: gene expression (log2); red line: Holstein; blue line: Jersey.



Supplementary Figure 2: a,b) The boxplots show the two significant eQTLs with the associated genes for high concentrate diet group analysis. x-axis: genotypes; y-axis: gene expression (log2); red line: Holstein; blue line: Jersey.